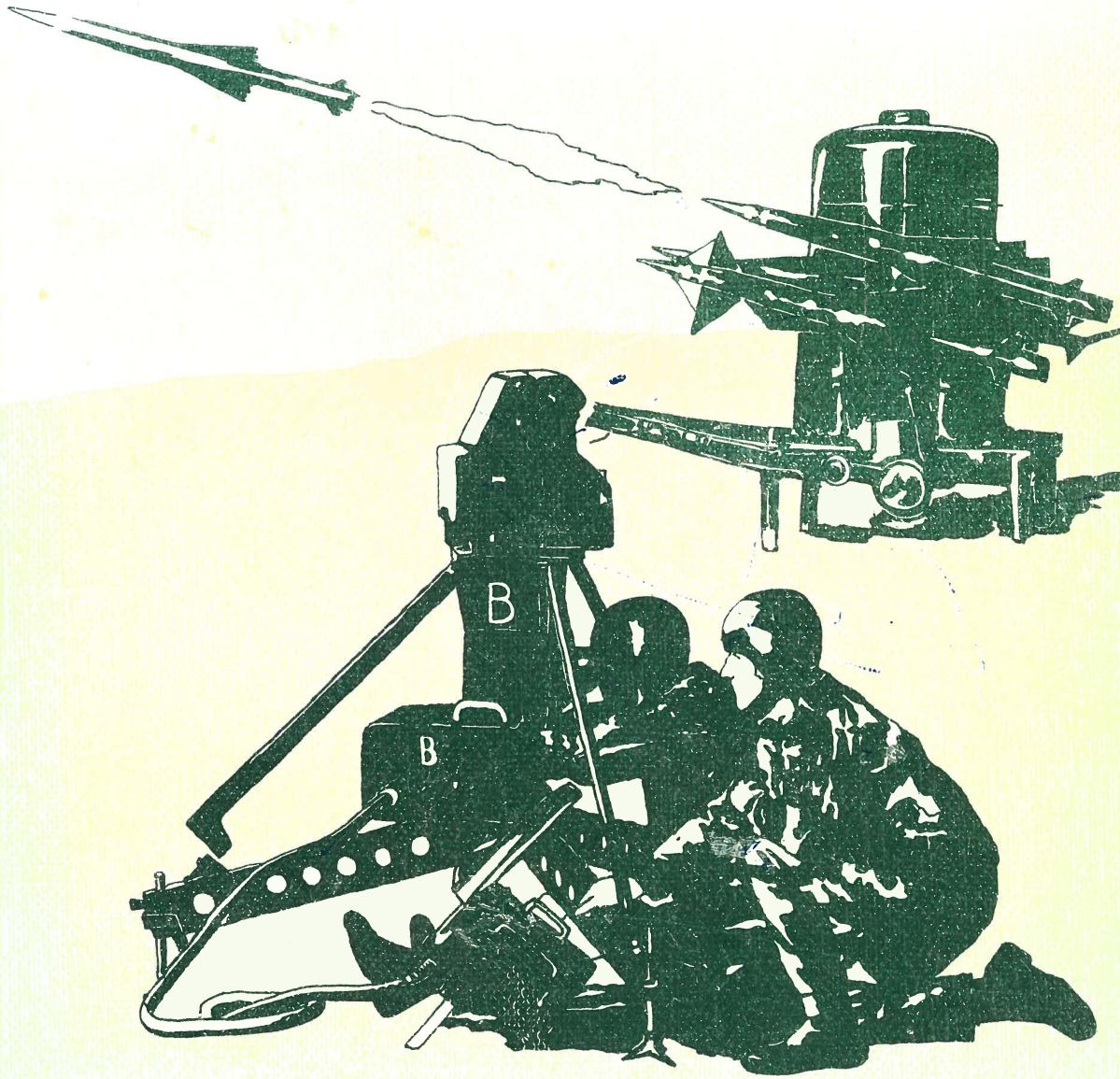
Guided Weapon Control Systems

P. Garnell & D.J. East

Royal Military College of Science, Shrivenham, Swindon, England



Pergamon Press

GUIDED WEAPON CONTROL SYSTEMS

BY

P. GARNELL
Principal Lecturer

AND

D. J. EAST
Senior Lecturer, Royal Military College of Science, Shrivenham, Swindon, England

PERGAMON PRESS
OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada Ltd., 75 The East Mall, Toronto, Ontario, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 Rue des Ecoles, 75240 Paris, Cedex 05, France
WEST GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Pferdstrasse 1, Frankfurt-am-Main, West Germany

---

First edition 1977

Library of Congress Cataloging in Publication Data

Garnell, P.

Guided weapon control systems.

Includes index.

1. Guided missiles–Control systems. I. East, D. J. joint author. II. Title.

UG1310.G37 1977 623.4'51 76-40061

ISBN 0-08-019691-8

In order to make this volume available as economically and rapidly as possible the author's typescript has been reproduced in its original form. This method unfortunately has its typographical limitations but it is hoped that they in no way distract the reader.

Printed in Great Britain by
Butler & Tanner Ltd, Frome and London

CONTENTS

CHAPTER 6 AUTOPILOT DESIGN

CHAPTER 7 LINE OF SIGHT GUIDANCE LOOPS

CHAPTER 8 HOMING HEADS AND SOME ASSOCIATED STABILITY PROBLEMS

CHAPTER 9 PROPORTIONAL NAVIGATION AND HOMING GUIDANCE LOOPS

CHAPTER 10 WIENER FILTER THEORY APPLIED TO GUIDANCE LOOP DESIGN

CHAPTER 11 MODERN CONTROL THEORY APPLIED TO GUIDANCE LOOP DESIGN

CHAPTER 12 KALMAN FILTERS

PREFACE

During the last twenty five years a large number of textbooks have been written on the subject of automatic control. Many of these books are introductory student texts concerned with basic theory and the examples used for illustration are often, of necessity, so simplified or contrived that the embryo control engineer gains little or no appreciation of the true nature of the problems that will be encountered in practice. The present authors believe that this text, which concentrates in depth on the use of closed loop control theory in one particularly rich and varied field of application, viz the design of guided weapon systems, will be a useful supplement to the aforementioned introductory texts and will provide further insight into the uses and limitations of the basic theory without being constrained to "examination" type problems and their solutions. We hope, therefore, that this book will be of interest to lecturers and postgraduate students taking control and systems orientated courses as well as to those engineers actively involved in guided weapon design who require a compact reference book directly related to their own field. The only essential prerequisite is a basic knowledge of classical control theory and familiarity with such terms as bandwidth, damping ratio, phase margin, steady state gain etc.

The subject matter of this book is based on lecture notes given to the Guided Weapon Systems (M.Sc.) Course at the Royal Military College of Science; this course is the only one of its type in this country and has been running continuously for twenty six years.

Finally, it is a pleasure to acknowledge the suggestions and discussions with our many colleagues, both from industry and the government establishments, and we regret that they are too numerous to mention individually. Also, we are grateful to Mrs A Hare who typed the manuscript and Mrs H Killeen and Mrs S Greener who prepared the drawings.

Shrivenham
August 1976

P Garnell
D J East

CHAPTER 1

THE PERFORMANCE OF TARGET TRACKERS

1.1 INTRODUCTION

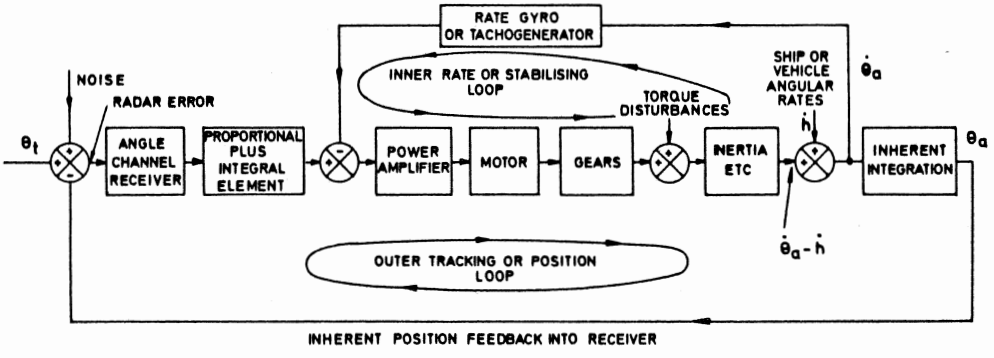
A guided missile is one which is usually fired in a direction approximately towards the target and subsequently receives steering commands from the guidance system to improve its accuracy. Inertial guidance is often used in medium and long range missiles (over 40 km say) when the intention is to hit a given map reference. The techniques used in such systems are quite different from those used in most short and medium range systems; moreover they have been adequately described elsewhere (1), (2), (3). The guidance-control systems covered in this book are command systems and homing systems. There is much in common between these two systems; for instance one has to track the target in both systems. In command systems the tracker is usually stationary or moving slowly (e.g. the target tracker could be on a ship). In homing systems the target tracker is in the missile and in such a case it is the relative movement of target and missile which is relevant. The special tracking problems associated with homing are considered in chapters 8 and 9; so in this chapter we assume that the tracker speed is small enough not to influence the kinematics of the engagement seriously.

1.2 A TRACKER SERVO

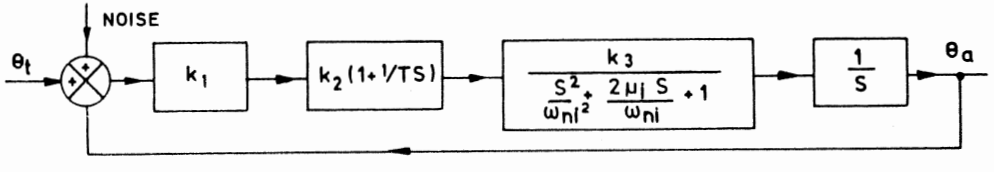
A target tracker attempts to align its electrical null axis (or "boresight" as it is often called) in elevation and azimuth with a line joining the tracker and target called the line of sight (LOS). There are two identical servo systems to do this, so only one will be considered. The first and essential requirement is for a device which produces two signals (one up-down and the other left-right) proportional to the misalignment between the LOS and the boresight. The angular error detecting mechanism associated with a tracker is either a radar receiver or an optical (and this includes infra-red) signal processing system (4).

Most of these error detectors are very non-linear for large misalignments but are regarded as essentially linear for small misalignments of about 1^{o} or less; tracking errors are rarely as large as this. The receiver may well provide range and range rate information as well but since we are concerned

only with angle measurement we will call this particular aspect of the receiver the "angle channel receiver". A typical target tracking servo loop is shown in Fig 1.2-1(a). The angle channel receiver produces a signal proportional to the misalignment between target and its own boresight, $\theta_t - \theta_a$. Since it is a linear device, and no reckonable time lag is associated with it, its transfer function is a simple gain k_1 volts/radian (misalignment). This error signal is fed to a proportional plus integral amplifier whose transfer function is $k_2\ (1 + 1/Ts)$. The usual servo components now follow, a power amplifier, motor (electric or hydraulic) and a gear speed reducer, together with the lumped inertia and viscous friction, if any. If the tracker is immobile some rate feedback is usually supplied by a tacho-generator. Typical transfer functions are shown in Fig 1.2-1(b). Provided sufficient power is available in the amplifier and motor, the effect of moderate



(a) Block diagram



(b) Transfer functions

FIG 1.2-1 Target tracker

or high gain in this rate loop is to reduce the lags in these components. If the tracker is on a moving vehicle, rate feedback is provided by a rate gyro (details in chapter 5). This has the additional effect of stabilising the antenna to a large degree against base motion; call this base or hull motion $\dot{h}$. Since the servo drives relative to the hull it will produce a speed relative to the hull $\dot{\theta}_a - \dot{h}$. If the open loop transfer function of this

inner loop is $K'\ G(s)$ where K' is the d.c. gain, in the absence of an error signal from the radar we can write

$$(\dot{\theta}_a - \dot{h}) = - K'\ G(s)\ \dot{\theta}_a$$

Rearranging this is

$$\frac{\dot{\theta}_a}{\dot{h}} = \frac{1}{1 + K'\ G(s)} \qquad (1.2\text{-}1)$$

Since $G(s) \to 1$ at low frequencies and $\to 0$ at high frequencies this equation is saying that if K' is large, say 100 or more, base motion isolation is very good at low frequencies. High gains are difficult to achieve in really heavy equipments due to the resilience of the gear train, but even tank gun and turret stabilisers using rate gyros achieve open loop gains well in excess of 100.

We now consider the closed loop transfer function of the whole servo. If there are two effective lags in the rate loop and its steady state gain is k_3 rad/sec/volt its response is completely defined by an undamped natural frequency ω_{ni} and a damping ratio μ_i. Since there is the usual inherent integration from antenna speed to position the closed loop transfer function is easily shown to be

$$\frac{\theta_a}{\theta_t} = \frac{Ts + 1}{\frac{s^4}{c^2\omega_{no}^4} + \frac{2\mu_i s^3}{c\omega_{no}^3} + \frac{s^2}{\omega_{no}^2} + Ts + 1} \qquad (1.2\text{-}2)$$

where $\omega_{no}^2 = k_1k_2k_3/T$ and represents the undamped natural frequency of the system if the rate loop lag is negligible and $c = \omega_{ni}/\omega_{no}$; $\mu_o = \sqrt{Tk_1k_2k_3}/2$. Since the coefficient of "s" in the numerator is the same as that in the denominator the system exhibits zero velocity lag, and has a steady state error to a constant input acceleration α of α/ω_{no}^2. This is easily seen when one visualises a constant error θ_ε. Since there is an integrator between this signal and the output speed the *slope* of the output speed is $\theta_\varepsilon k_1k_2k_3/T = \theta_\varepsilon\ \omega_{no}^2$; and the slope of the speed output is the acceleration α.

1.3 TRACKING ACCURACY IN THE ABSENCE OF NOISE

In order to define the performance of any system one must first define the input. In practice one specifies a maximum speed of the target, a minimum tracking range and possibly a minimum "crossing range". This latter range is defined, on the assumption that the target continues to fly straight, as the shortest distance that it can be from the tracker. When it is at this

point it is "passing" the tracker and is at the "point of closest approach". It is important to note that "crossing range" means slant range. If the point of closest approach in ground range is 4 km and the target height is 3 km the crossing range is 5 km. A target may well manoeuvre for a short time before attacking but during most or all of the time targets fly straight with constant velocity. We will therefore take such a target as the standard; surprisingly enough moderate target manoeuvres do not add greatly to the tracking task. In Fig 1.3-1 the crossing range is d and the corresponding ground range is d'. The instantaneous slant range is r and the corresponding ground range is r'.

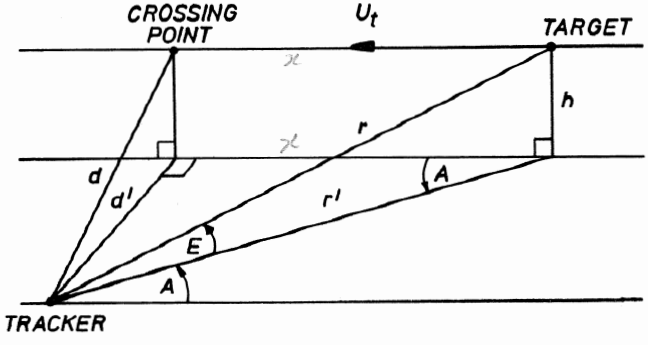


FIG 1.3-1 Angles and distances associated with tracking a target flying at constant altitude

In the following analysis it is assumed that the target speed U_t is constant.

$$\dot{A} = U_t \sin A/r' = U_t \sin^2 A/d' \qquad (1.3\text{-}1)$$

$$\ddot{A} = \frac{d}{dA} \cdot \frac{dA}{dt} \cdot \frac{dA}{dt} = \frac{2U_t}{d'} \cos A \sin A \frac{U_t \sin^2 A}{d'}$$

$$= \frac{U_t^2}{(d')^2} \sin 2A \sin^2 A = \frac{U_t^2 \sin 2A}{(r')^2} = \frac{U_t^2 \sin 2A}{r^2 \cos^2 E} \qquad (1.3\text{-}2)$$

Hence if R_{min} is the minimum specified slant range the maximum angular acceleration for a given E occurs when $A = 45^o$ and is given by

$$\ddot{A}_{max} = U_t^2/(R_{min}^2 \cos^2 E) \qquad (1.3\text{-}3)$$

It should be noted that angular accelerations can be very large in azimuth if the angle of elevation is large; in practice the maximum angle E is often limited to about 55^o.

Similarly, expressions for $\dot{E}$ and $\ddot{E}$ can be obtained:

$$\dot{E} = \frac{U_t}{r} \sin E \sin A \qquad (1.3\text{-}4)$$

$$\text{and } \ddot{E} = -\frac{U_t^2}{r^2} \tan E \{1 - \cos^2 A (1 + 2\cos^2 E)\} \qquad (1.3\text{-}5)$$

If now $A = 90^0$ (i.e. target is crossing)

$$\ddot{E}_{max} = -\frac{U_t^2}{R_{min}^2} \tan E \qquad (1.3\text{-}6)$$

and if $A = 0^0$ (i.e. target directly ahead)

$$\ddot{E}_{max} = \frac{U_t^2}{R_{min}^2} \sin 2E \qquad (1.3\text{-}7)$$

It is seen therefore that, even for the apparently straightforward case of a constant speed target flying at constant height, angular rates and accelerations are not constant. We can if necessary obtain expressions for the third and higher derivatives. Had the angular acceleration been *constant* we could have calculated the constant following error by the method discussed in section 1.2; in practice the tracking error will vary with time. The graphs drawn in Figs 1.3-2 and 1.3-3 show the ratio of the actual following error θ_e to the approximate value given by the instantaneous angular acceleration divided by the loop gain. The graphs are applicable to motion in azimuth when the angle of elevation is small or to motion in elevation when A is small; the generalised angle θ has been used instead of E or A since the results are then applicable to both channels. It is seen that when the acceleration is increasing ($\theta < 60^0$) the following error is less than the approximate value, and when it is decreasing the error is greater. This is explained when one remembers that displacements are not accelerations and that a following error in *position* takes time to integrate up. When one takes the rather more complicated inputs (e.g. motion in azimuth when the angle of elevation is not small) one arrives at a similar conclusion: the actual following error to a good first approximation for a type 2 servo is the *actual* input acceleration divided by the loop gain. This assumption is made in the following design example.

Suppose therefore that the problem is to track targets flying up to 600 m/sec as accurately as possible at ranges between 4 and 32 km. The design aim is, under the worst conditions, for the r.m.s. tracking error not to exceed 0.3 milliradians. If we concern ourselves with the elevation channel only and assume the tracking elevation will not exceed 45^0, equation 1.3-6 or 1.3-7 can be used to compute $\ddot{E}$. This occurs when $E = 45^0$ and is given by

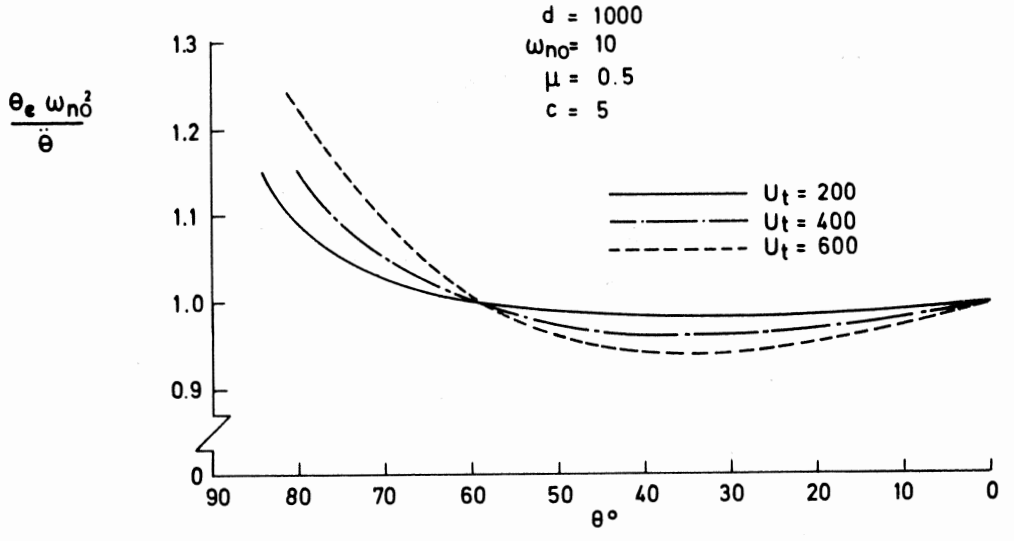
$$\ddot{E}_{max} = 600^2/16 \times 10^6 = 2.25 \times 10^{-2} \text{ rad/sec}^2$$

The minimum open loop gain for θ_e to be 3×10^{-4} radians is $2.25 \times 10^2/3 = 75$ rad/sec^2/rad. The minimum outer loop natural frequency ω_{no} is therefore $\sqrt{75} = 8.7$ rad/sec.

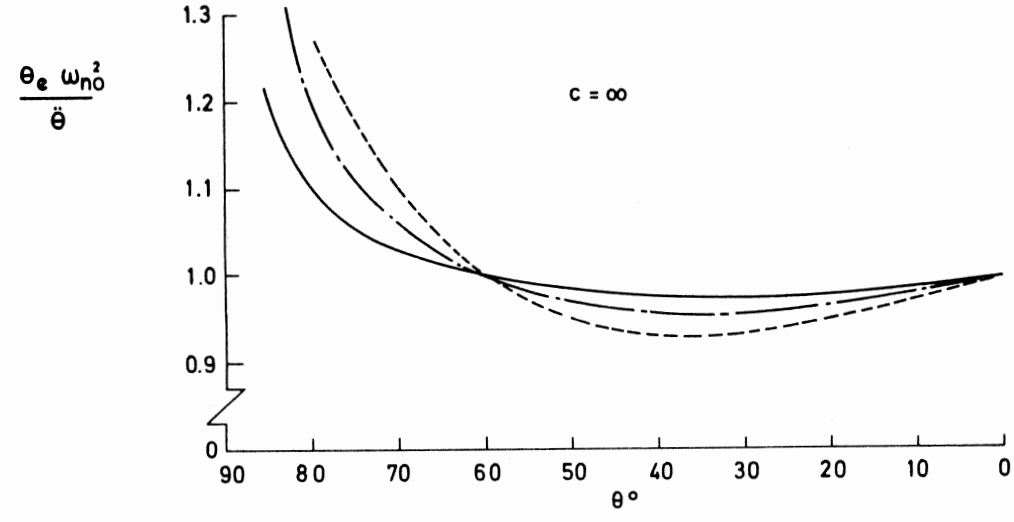
The damping ratios of the outer and inner loops and the ratio of the bandwidths of these two loops have little effect on the tracking accuracy, but they will, of course, affect the stability margin of the system. Some indication of the optimum values for these parameters can be obtained by considering the performance in the presence of noise.

1.4 THE EFFECT OF THERMAL NOISE

Noise is present in all receivers and the reader is referred to Skolnik (4) for a discussion on the main sources of noise in radar receivers. Our problem stems from the fact that in closed loop systems high amplification of error signals is needed to obtain good system accuracy; nevertheless amplification of signals amplifies noise as well, so that some form of compromise is necessary. The main source of noise in radar receivers is "thermal noise" because electrons in any conductor at a temperature other than absolute zero are always in random motion. This motion gives rise to an electrical noise voltage which is essentially "white" i.e. its spectrum is independent of frequency from d.c. to a frequency far in excess of any servo tracker bandwidth. There are many other sources of noise associated with receivers including environmental background noise but in practice it is found that if receiver noise is significant it is largely due to thermal noise, and therefore is sensibly constant for a given receiver. The actual noise output expressed as a mean square voltage however is not constant. If the incoming signal is strong (e.g. a large target at short range) an automatic gain control reduces the gain of the IF amplifier, in order to keep the output independent of the signal power and this effectively reduces the noise output. The result is that the signal-to-noise ratio varies. If one is illuminating and tracking a target the received power will vary inversely as the (range)^4, all other things being equal. An assumption that is usually made therefore is that the mean squared noise output from a receiver for a given target, range and atmospheric conditions varies inversely as the signal-to-noise ratio and is proportional to (range)^4. Since all angle channel receivers are designed to produce a voltage proportional to the angular misalignment it follows that the mean square noise output can be regarded as a mean square angle. Hence, thermal noise is often referred to

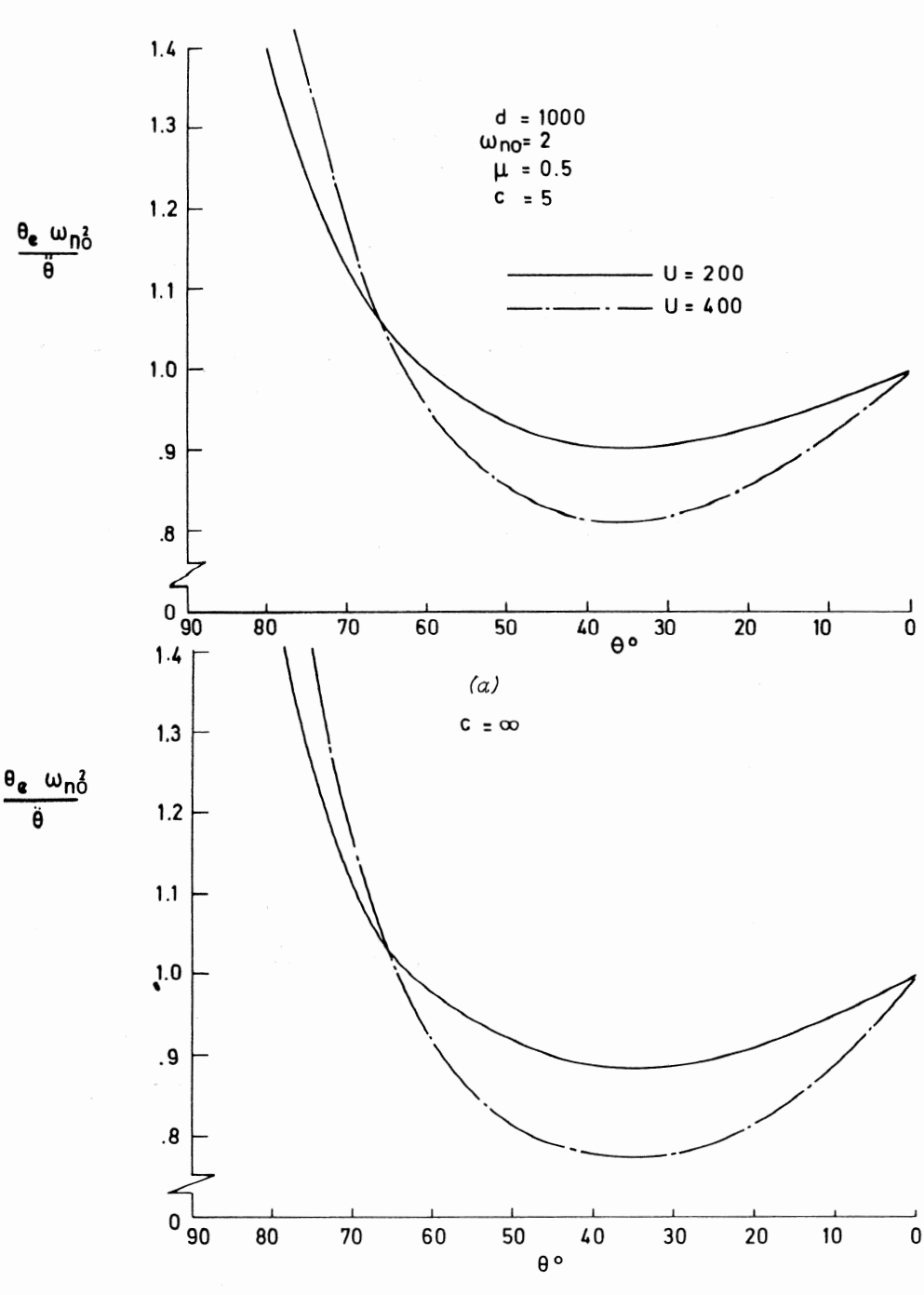


(a)



(b)

FIG 1.3-2 Following error for a type 2 servo



(b)

FIG 1.3-3 Following error for a type 2 servo

as "angular" noise. Let the spectrum of this noise be defined as

$$\Phi_a(\omega) = K_a^2 \text{ rad}^2/\text{rad/sec}, \text{ where } K_a \text{ is a constant.}$$

The problem now is to determine the mean square value (and hence r.m.s. value) of the antenna "jitter" due to this noise. It is most important to note at this stage that, ideally we require a servo to have zero bandwidth so that all this noise is filtered out. Consider now the mean square output σ^2 of a filter whose transfer function in frequency form is $1/1 + j\omega T$. The input is white noise whose spectrum is constant up to a frequency ω_b. The mean square output is therefore

$$\sigma^2 = K_a^2 \int_o^{\omega_b} \frac{d\omega}{(\sqrt{1 + \omega^2 T^2})^2} = K_a^2 \int_o^{\omega_b} \frac{d\omega}{1 + \omega^2 T^2}$$

$$= \frac{K_a^2}{T} \tan^{-1} \omega_b T$$

Provided $\omega_b T > 10$ say this approximates to $\pi K_a^2/2T$. This means therefore that the mean square output would be the same as that from another filter whose pass band was flat up to a frequency equal to $\pi/2T$ and then cut off completely at this frequency. In other words a simple first order lag has an effective bandwidth to white noise of $\pi/2$ times its own bandwidth of $1/T$. It is easily shown that the effective noise bandwidth of any linear filter is a function of the coefficients of the transfer function. In general a linear system will have a transfer function of the form

$$\frac{b_m s^m \ldots\ldots + b_1 s + b_o}{a_n s^n \ldots\ldots + a_1 s + a_o} = \frac{b(s)}{a(s)}$$

where the order of the numerator is at least one less than that of the denominator. Integrals of the form

$$I = \int_o^{\infty} \frac{|b(j\omega)|^2 \, d\omega}{|a(j\omega)|^2}$$

can be evaluated and Table 1.4-1 gives values for systems up to fourth order assuming as we always can that $a_o = 1$. We justify integrating to $\omega = \infty$ because the noise bandwidth of thermal noise is so high.

TABLE 1.4-1

Order of system	Value of I
First order	$\frac{\pi b_o^2}{2a_1}$
Second order	$\frac{\pi}{2}\,\frac{\frac{b_1^2}{a_2}+b_o^2}{a_1}$
Third order	$\frac{\pi}{2}\,\frac{\frac{b_2^2 a_1}{a_3}+b_1^2-2b_ob_2+b_o^2a_2}{a_2a_1-a_3}$
Fourth order	$\frac{\pi}{2}\,\frac{b_3^2(\frac{a_1a_2}{a_3a_4}-\frac{1}{a_4})+(b_2^2-2b_1b_3)\frac{a_1}{a_3}+b_1^2-2b_ob_2+b_o^2(a_2-\frac{a_1a_4}{a_3})}{a_1a_2-\frac{a_1^2a_4}{a_3}-a_3}$

It is instructive to evaluate the effective noise bandwidth of a simple unit gain second order system where $b_1 = 0$, $b_o = 1$, and $a_1 = 2\mu/\omega_n$. The equivalent noise bandwidth is therefore

$$\frac{\pi}{2}\frac{\omega_n}{2\mu} = \frac{\pi}{4}\frac{\omega_n}{\mu}$$

If $\mu = 1$, a second order system is equivalent to two first order systems in series with the same time constant equal to $T = 1/\omega_n$. The effect of these "two" filters is to halve the effective noise bandwidth. Nevertheless a low damping ratio can increase the noise bandwidth very considerably and this is clearly due to some frequency components being amplified. Now consider the simplest model of our type 2 tracking servo with a negligible lag in the rate loop $(\sigma \to \infty)$ i.e. the transfer function is

$$\frac{b_1 s + 1}{a_2 s^2 + a_1 s + 1}$$

where $b_1 = a_1 = 2\mu_o/\omega_{no}$ and $a_2 = 1/\omega_{no}^2$. The equivalent noise bandwidth is

$$\frac{\pi}{2}\,\omega_{no}\,(2\mu_o + 1/2\mu_o)$$

If this is differentiated with respect to μ_o and equated to zero we find that a minimum occurs with $\mu_o = 0.5$. The minimum noise bandwidth is therefore $\pi\omega_{no}$

which is double that for a simple second order system with the same damping ratio. We note that a proportional plus integrating element with a transfer function of $1 + 1/Ts$ is equivalent to $(1 + Ts)/Ts$ and is therefore equivalent to pure integration with a gain of $1/T$ in series with a proportional plus derivative element. Excessive feed-forward of derivative of the error has the effect of increasing the bandwidth whereas feedback of derivative of the output in general has the effect of reducing the bandwidth. For our original system with a reckonable lag in the rate loop the equivalent noise bandwidth from Table 1.4-1 is

$$\frac{\pi\omega_{no}\ (\mu_o - 1/4c\mu_1 + 1/4\mu_o)}{1 - \mu_o/\mu_1 c - \mu_1/c\mu_o}$$

What are the optimum values for μ_o, μ_i and c for minimum noise bandwidth? The reader may wish to verify that $\mu_o = 0.5$ is nearly always near-optimum, and that c should be at least 4. The bandwidth is rather insensitive to the damping ratio of the inner loop provided c is 4 or more, but $\mu_i = 0.5$ is probably the best choice if c or μ_o are going to vary. The reason why a low value of c causes such a large increase in the equivalent noise bandwidth is that a slow inner loop erodes the stability margin of the overall system. Fig 1.4-1 shows the equivalent noise bandwidth as a function of μ_o and c with μ_i set to 0.5.

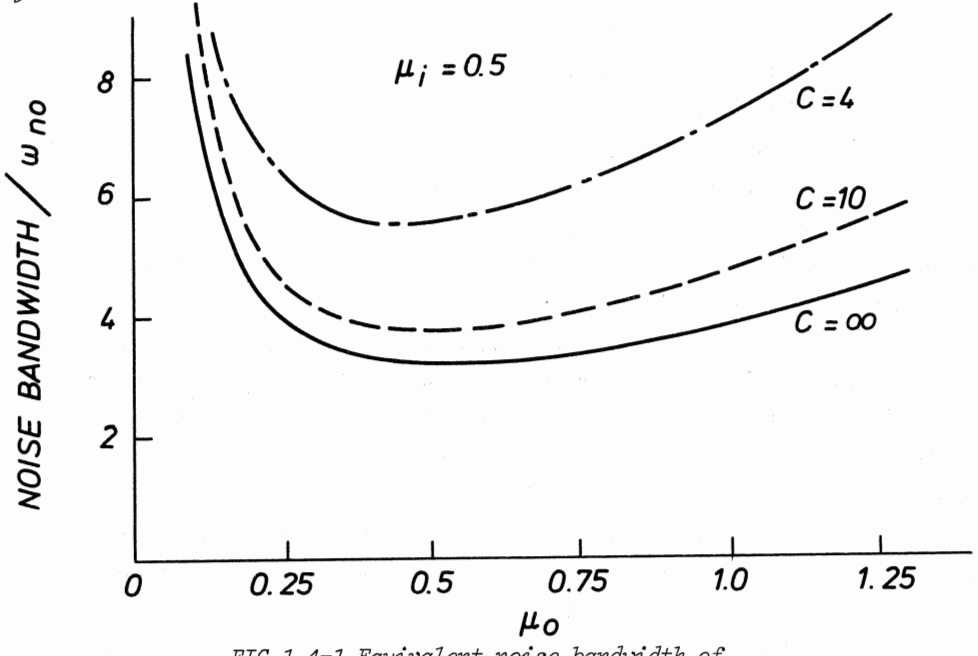


FIG 1.4-1 Equivalent noise bandwidth of a fourth order type 2 servo

Let us therefore now fix the main design parameters for the servo. $\omega_{no} = 10$ gives us the necessary tracking accuracy with a small margin. $\mu_o = \mu_i = 0.5$ and $c = 5$ say to minimise the equivalent noise bandwidth. Any compensators such as lag-lead and lead-lag networks are placed in the rate loop to make possible a fast inner loop, the faster the better. The equivalent noise bandwidth is therefore

$$\pi \omega_{no} \frac{(0.5 - 0.1 + 0.5)}{1 - 0.2 - 0.2} = 47 \text{ rad/sec}$$

We now consider a particular antenna and receiver design and assume an equivalent echoing area of target. The spectral density of the thermal noise (to take a realistic figure) is given by

$$\Phi_a = K_a^2 = 8.0 \times 10^{-15} R_t^4/R_o^4 \text{ rad}^2/\text{rad/sec}$$

where R_o is a reference range of 1 km and R_t is the actual range of the target in km. The mean square jitter of the antenna due to thermal noise is therefore

$$\sigma_n^2 = 8.0 \times 10^{-15} \times 47 \times R_t^4 \text{ rad}^2$$

The r.m.s. value σ_n is

$$\sigma_n = 6.13 R_t^2/10^7 \text{ rad}$$

which is 0.01×10^{-3} rad at 4 km
0.04×10^{-3} rad at 8 km
0.16×10^{-3} rad at 16 km
0.35×10^{-3} rad at 24 km
0.63×10^{-3} rad at 32 km

It is clear that we cannot meet the requirement of an accuracy of 0.3×10^{-3} rad at the longer ranges due to excessive servo bandwidth; but a large servo bandwidth is not required at these ranges. One method therefore of improving the overall performance is to reduce the gain of the loop starting at ranges in excess of 16 km say. To keep the damping ratio of the outer loop constant the values of k_2 and T, see Fig 1.2-1(b) would have to be varied together. This is easily implemented in practice as range information will almost certainly be available from the receiver. This will not only decrease the bandwidth of the outer loop but will also increase the value of c thus reducing the effective noise bandwidth still further. Table 1.4-2 shows the estimated tracking accuracy using a fixed gain servo loop compared with a variable gain loop. Since there is no correlation between the acceleration of the sight line and thermal noise in the receiver and that the noise output of the receiver will have a normal distribution we are justified in

assuming

$$\sigma_t = \sqrt{\sigma_a^2 + \sigma_n^2}$$

where σ_t is the total r.m.s. tracking error and σ_a is the tracking error due to sight line motion.

TABLE 1.4-2 TRACKING ERRORS IN MILLIRADIANS

RANGE km	FIXED GAIN			VARIABLE GAIN			
	σ_a	σ_n	σ_t	ω_{no}	σ_a	σ_n	σ_t
4 km	0.225	0.01	0.225	10	0.225	0.01	0.225
8 km	0.056	0.04	0.069	10	0.056	0.04	0.069
16 km	0.014	0.16	0.16	5	0.056	0.09	0.11
24 km	0.006	0.35	0.35	2.5	0.10	0.14	0.17
32 km	0.003	0.63	0.63	1.4	0.175	0.19	0.26

1.5 A SELF OPTIMISING SERVO

We are not suggesting that the above servo with a variable gain is an "optimum" servo. For instance one can argue that the gain at the lowest ranges should be increased to increase the overall accuracy. One could argue that the structure of the servo loop is not the "best"; a lag-lead compensator could be used to increase the gain, or one might even envisage a type 3 servo. However, there are objections which are far more fundamental than these. We have designed for the worst possible value of θ and in most engagements a value of 45^o will not be encountered. We have assumed a given equivalent reflecting area of the target and we have assumed no jamming noise; and we have assumed a given speed of target. Targets could be much slower than this or conceivably faster. Atmospheric conditions (such as rain) and a smaller or larger target could easily result in the assumed noise level to be seriously in error.

All these arguments are suggesting that in order to be truly adaptive we ought to make some intelligent enquiries during the engagement in order to find out just what is going on. An obvious method would be to fit angular accelerometers to the radar antenna but in view of the low angular accelerations involved this method could be rather costly. A simple method would be to make use of the information already available i.e. the tracking error signal. If there is no noise present this signal will essentially be a "d.c." signal as it will be changing very slowly with time. There will certainly be no high frequency components present. Now consider the nature of the "error" signal when there is white noise present from the receiver. If this signal is fed

to a filter with a pass-band from say 10 Hz to 1 kHz, the filtered output will know virtually nothing of the true tracking error. A large output will merely indicate that there is a lot of noise in the system. To obtain more precise information concerning errors due to accelerating sight lines we need to pass the error signal through a low-pass filter with a pass-band from 0-2 rad/sec say. If the white noise has a bandwidth of 5 kHz or thereabouts, the output from this filter knows nothing of the noise in the system. Fig 1.5-1 is suggesting that the outputs from these filters are

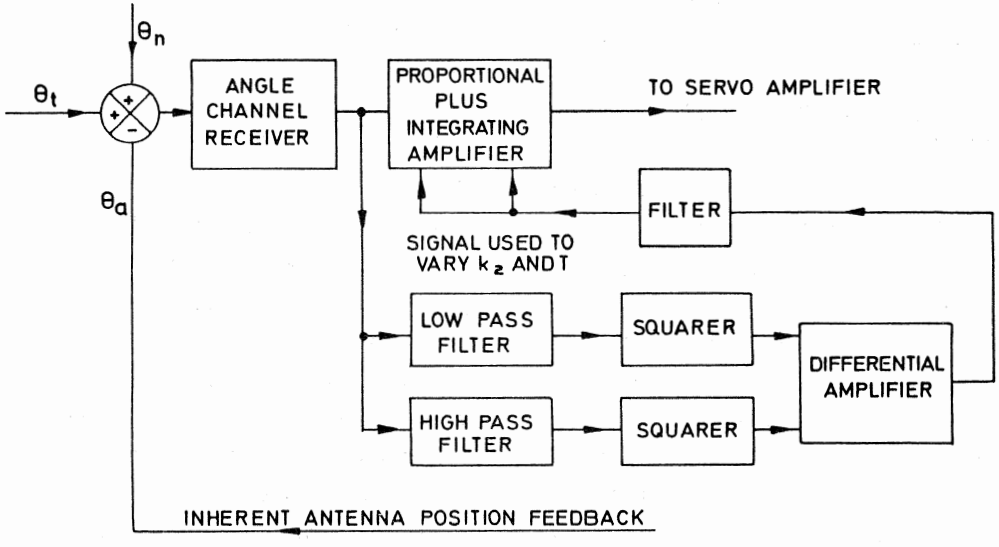


FIG 1.5-1 The use of filters in a self-optimising servo

both squared (to make them invariably positive) and the respective outputs fed to a differential amplifier. If this difference is then fed to a filter (say an integrator) the output of this filter is used to vary the gain and integrating time constant of the proportional plus integral element. If the output from the low pass filter is less than that of the high pass filter then the arrangement must be to decrease the gain of the system and vice-versa. If the gains of these filters are chosen correctly the tracking error can be minimised. Since the tracking error due to angular acceleration is inversely proportional to $(\text{bandwidth})^2$ and the jitter is proportional to $(\text{bandwidth})^{\frac{1}{2}}$, it is easily shown that the system is optimised when the mean square jitter due to white noise is four times the mean square true tracking error.

1.6 THE EFFECT OF OTHER INPUTS AND DISTURBANCES

An important contribution to tracking accuracy when using radar is the phenomenon of "glint". It is well known that the apparent radar centre of a target (other than a perfect sphere) moves during flight due to target motion causing differential phase changes at the receiver between the returns scattered from different parts of the target. Such motions are caused by vibration, target manoeuvres, range closure causing a changing target aspect and air turbulence (aircraft) and sea waves (ships). Sometimes these angular motions of the target cause the radar to interpret the true centre of area to be well outside the target outline. The fluctuation rate is proportional to target angular rate and also to microwave frequency. In strike aircraft glint is largely a yaw phenomenon and its r.m.s. value is often of the order a fifth of the wingspan and this means that its r.m.s. value could be in the range 2-4 metres. These figures are very approximate as the amount of glint can vary considerably with target aspect and will vary from target to target depending on the "smoothness" of the reflecting area (5). Its spectrum approximates well to white noise passed through a first order lag of time constant T_g i.e.

$$\Phi_g(\omega) = \frac{K_g^2}{1 + \omega^2 T_g^2} \, m^2/\text{rad/sec} \qquad (1.6\text{-}1)$$

$$\text{where } L_g^2 = \pi K_g^2 / 2T_g \qquad (1.6\text{-}2)$$

and is the mean square value of the glint and T_g is typically in the range 0.1 to 0.25 secs.

This means that glint is a low frequency source of noise and the energy will be largely within the pass-band of the radar servo, except perhaps at the longer ranges where the servo bandwidth is so low. It is instructive to calculate the r.m.s. antenna movement due to glint for different servo bandwidths. Assume for simplicity that the servo approximates to a second order system i.e. one with a very fast inner rate loop. The situation is depicted in Fig 1.6-1.

$$G_1(s)\, G_2(s) = \frac{1}{1 + T_g s} \cdot \frac{2\mu_o s/\omega_{no} + 1}{s^2/\omega_{no}^2 + 2\mu_o s/\omega_{no} + 1}$$

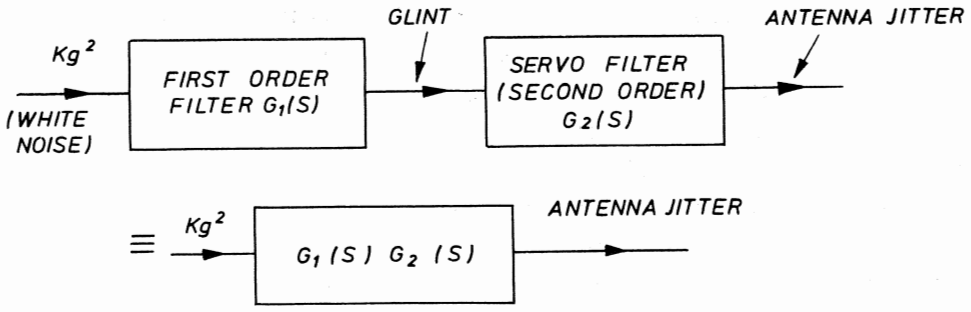


FIG 1.6-1 Glint represented as white noise having passed through a first order filter

If $\mu_o = 0.5$ as before, and using Table 1.4-1 we find that the equivalent noise bandwidth is

$$\frac{\pi}{2} \cdot \frac{b_1^2 + a_2}{a_2 a_1 - a_3} = \frac{\pi}{2T_g} \cdot \frac{2 + T_g\omega_{no}}{T_g\omega_{no} + 1 + 1/\omega_{no}T_g}$$

The mean square output or jitter due to glint is therefore

$$\sigma_g^2 = K_g^2 \cdot \frac{\pi}{2T_g} \cdot \frac{2 + T_g\omega_{no}}{T_g\omega_{no} + 1 + 1/\omega_{no}T_g}$$

$$= L_g^2 \cdot \frac{2 + T_g\omega_{no}}{T_g\omega_{no} + 1 + 1/\omega_{no}T_g} \qquad (1.6\text{-}3)$$

Table 1.6-1 shows the ratio of r.m.s. antenna jitter to the r.m.s. value of the glint for various normalised servo bandwidths.

TABLE 1.6-1 ANTENNA JITTER DUE TO GLINT

$\omega_{no}T_g$	0.25	0.5	1	2	10	∞
σ_g/L_g	0.65	0.85	1.00	1.07	1.04	1

This means that in general the r.m.s. jitter due to glint is about the same value as the r.m.s. value of the glint itself, except when the undamped natural frequency of the servo is less than about one half that of the glint bandwidth. It is pointed out that the jitter has been expressed in metres rather than in angular measure. This is because glint is essentially an apparent movement of the target in metres. The radar sees this as an angle but this will result in a movement of the tracking beam in metres at the target. An r.m.s. jitter of 2m at a range of 4 km is an angular tracking error of 0.5×10^{-3} rad due to this cause alone and this means that we cannot meet the specification at the shorter ranges.

There is however a well-known technique called frequency agility which can be used with monopulse radars for spreading the glint spectrum over a wider frequency band (6), (7). The effect of changing the transmitted frequency is to change the coherent interference pattern at the aerial of the waves scattered from the target by altering the phase differences between the contributions from the individual scatterers. Frequency agility however, does not alter the r.m.s. value of the glint. If, for example, the effective bandwidth can be increased by a factor of eight and the original time constant is 0.16 secs then the apparent glint bandwidth now extends to 50 rad/sec, and much of this energy will be filtered out by the servo. If, however a reckonable amount of very low frequency glint noise did appear at the receiver output its effect, in the self-optimising servo just described, would be to increase the bandwidth of the servo as it thinks this "error" is a true tracking error. A possible safeguard would be to place a modest limit on the maximum allowable loop gain.

How does a target manoeuvre affect tracking accuracy? Suppose a target should suddenly make an evasive manoeuvre of 4g. The maximum effect on sight line acceleration will be if the flight path is down the line of sight such that the target acceleration is perpendicular to it. At 4 km range the angular acceleration of the sight line due to this manoeuvre is $4 \times 9.81/4000$ rad/sec^2 and this is approximately 1×10^{-2} rad/sec^2. This is only 40% of the maximum angular acceleration due to a straight flying target and therefore will not seriously affect the tracking accuracy. How does one assess errors due to amplifier biases, gear imperfections and random variations in friction? Fortunately nowadays d.c. amplifiers are readily available with offsets that can be trimmed easily and whose drift rates are negligible. If a large gear reduction is used between motor and load it will probably be difficult to design a high-gain rate loop because of the resilience of the drive; this introduces a very lightly damped high-order dynamic lag in the loop. Many trackers designed for high accuracy use high-torque low-speed electric drive motors driving direct or through a single gear reduction. These motors do not have a conventional rolling pin shaped armature; rather, the machine is pancake shaped. A direct drive eliminates all backlash also, and another contribution to instability is removed. Roughness of drive better than one arc second is claimed for some of these direct drive systems. Most modern tracking servos can keep servo noise down to less than 0.1 milliradian r.m.s.

REFERENCES

1. MCCLURE C.L. Theory of inertial guidance. Prentice Hall International Inc 1960.

2. SAVANT et al. Principles of inertial navigation. McGraw Hill Book Coy Inc 1961.

3. PITMAN G.R. Inertial guidance. John Wiley and Sons 1962.

4. SKOLNIK M.I. Introduction to radar systems. McGraw Hill Book Coy Inc 1962.

5. DELANO R. A theory of target glint or angular scintillation in radar tracking. Proc IRE, pp 1778-1784, December 1953.

6. LIND Göran. Reduction of radar tracking errors with frequency agility. IEEE Transactions on Aerospace and Electronic Systems Vol AES-4 No 3 May 1968.

7. BIRKEMEIER W.P. and WALLACE N.D. Radar tracking accuracy improvement by means of pulse-to-pulse frequency modulation. IEEE Trans Communications and Electronics January 1963.

CHAPTER 2

MISSILE SERVOS

2.1 SERVO REQUIREMENTS

Anything other than the warhead and fuze of a missile is dead weight. Minimum weight and volume are almost over-riding requirements for any missile servo. Good shelf life, low cost and reliability are obvious requirements, and a facility to test is an advantage although it can be argued that if reliability is excellent then a requirement to test, which involves time, training and test equipment, should be dispensed with. Servos for large radars and launchers are either electro-hydraulic or all-electric. New equipments of this type tend to use d.c. motors with constant field excitation and solid-state controlled armatures. Now a tactical guided missile can weigh anything from 10 kg to 1000 kg with flight times of 5 to 100 seconds or more. Some systems are liable to experience very noisy guidance signals while others may have a comparatively quiet ride. It is this diversity in size, duration of flight and type of duty cycle that has resulted in a number of successful designs of servos to operate control surfaces or thrust vector elements. Quite clearly a servo specification will call for a minimum dynamic performance and this could be conveniently expressed as "the phase lag in the frequency range a to b rad/sec shall not be worse than that of a quadratic lag of undamped natural frequency c rad/sec and damping ratio d". Some control over stability can be obtained by calling for a maximum value of closed loop amplitude ratio, or resonant gain as it is sometimes called. Now all missile servos incorporate mechanical stops to limit the angular travel to a safe limit, typically $\pm$ 15^{o}-25^{o}. Suppose $\hat{\theta}_o$ = 0.3 rad and we call for a performance no worse than that defined by ω_n = 150 and μ = 0.5. Since the maximum fin velocity is $\pm \omega \hat{\theta}_o$ = 45 rad/sec or about 450 rev/min one has to consider whether this comparatively large amplitude will ever be associated with this high frequency. It is argued in section 6.16 that fin rate saturation can have serious consequences in that it tends to unstabilise the autopilot. Since saturation, if it occurs, will be mainly due to low frequency glint noise and broad-band white thermal noise, it follows that an accurate estimate of noise levels is desirable before a servo specification is written. Hence, maximum fin rate will always be defined. The load inertia will be quoted of course in addition to the maximum aerodynamic hinge moment due to the control surface

centre of pressure not always coinciding with the axis of rotation. Some thrust vector methods involve considerable coulomb friction due to seals.

2.2 STORED COLD GAS SERVOS

The advantage of using a fluid operated motor compared with an electric device to produce a force or torque is the much higher pressures that can be utilised with high pressure gas or hydraulic fluid. In a stored gas system working pressures are typically about 5 MN/m^2 (1 $MN/m^2 \equiv 145$ lbf/in^2) and this is much higher than the effective pressure obtainable in any electric motor; this makes for a small final control element or "actuator". Fig 2.2-1 shows the main features of a pneumatic fin servo; the error signal is obtained by summing the input demand voltage with the negative feedback from a potentiometer. Compensation and amplification of the error signal are also performed electrically. These are conventional techniques and apply to all the servos to be described. Actuation power is obtained from stored gas which is released just before firing by means of a solenoid-operated start valve. A reducing valve regulates the pressure to a lower value such that the servo supply pressure remains approximately constant until the bottle pressure is less than the reduced pressure. The system illustrated represents a gas operated servo in its simplest form. The double acting actuator is the "area-half-area" type and operates in the on-off mode.

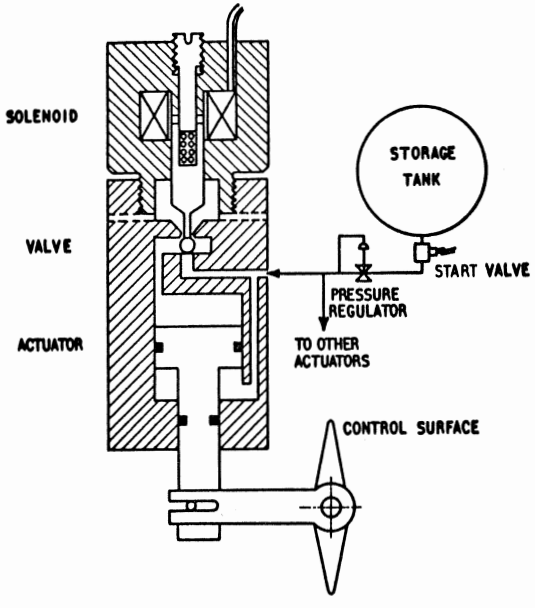


FIG 2.2-1 Mechanical and electro-mechanical components of a stored gas system

If the compensated error signal is converted into a pulse-width modulated signal the servo operates as a linear system with a superimposed high frequency carrier equal to the modulating frequency. If the compensated error signal is simply used to operate the solenoid when this signal is positive (say) and not to operate it when it is negative then a limit cycle will result. A small

high frequency limit cycle is not a bad thing in a servo where the coulomb friction is appreciable as it tends to improve the accuracy. The American firm Chandler Evans Inc has manufactured a number of successful guided weapon and aerospace pneumatic flight control systems (1). Helium or nitrogen is stored at 50-55 MN/m^2, and an electro-explosive cutter which ruptures a sealing membrance is a standard technique. Gas containers are subjected to a weight check and mass spectrometer leakage test. Such a system is attractive from a long term storage point of view and the reliability is excellent because the fluid is clean. The electronics and valve design are simple and cheap to manufacture.

If however the load inertia is considerable, the bandwidth obtainable with a pneumatic servo is limited; the dynamic lag from valve position to load speed is a second order one due to the compressibility of the gas and the pneumatic natural frequency decreases with an increase in load inertia. The pneumatic open loop frequency can be increased by increasing the gas supply pressure or by increasing the diameter of the actuator (2). An important design factor in any high performance pneumatic servo is an amplifier specifically designed to reduce the electrical time constant associated with the inductive solenoid winding (3). Pneumatic servos are usually associated with small missiles and fairly short times of flight; the weight of fuel plus bottle tends to be prohibitive if the total energy demand is large.

2.3 HOT GAS SERVOS

The advantage of burning cordite or some other mono-fuel such as iso-propyl nitrate and using the hot gas to drive the actuator is the reduction of the size and hence weight of the fuel container. The cordite can burn at the supply pressure and need not be stored at say ten times the supply pressure. Since however cordite burns at a greater rate at a higher pressure a relief valve must be used to allow generated gas to escape to atmosphere at times when the demand is low. Fig 2.3-1 shows a typical arrangement using an equal area double-acting piston. An area-half-area design is often favoured for hot gas servos, in which case only one valve is required per actuator; it is necessarily bigger however since the volume of gas to be supplied is larger. However, reliability is increased and manufacturing costs are reduced. Although a direct bleed from the main propulsion motor plus a heat sink has been used on a system with a flight time about 12 seconds, a separate charge of much cooler burning cordite is generally used. Since cordite produces very dirty hot gas

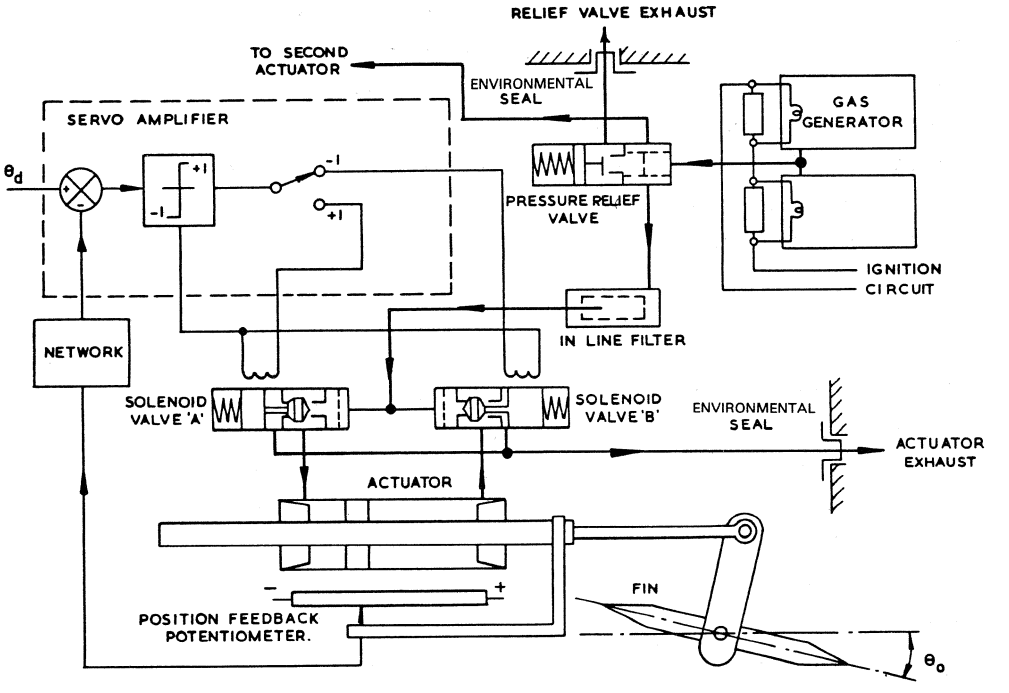


FIG 2.3-1 A hot gas servo

it will be appreciated that a long flight time creates some severe design problems, and most hot gas systems used so far have been restricted to missiles with short times of flight; but there seems to be no fundamental reason why this should always be so.

2.4 RAM AIR SERVOS

The concept of obtaining one's supply of air direct from the atmosphere has always been attractive. Although prototypes have been designed and successfully tested in a wind tunnel no such system in a production missile is known to the author. Fig 2.4-1 shows such a system, air being supplied through a number of pitot intakes positioned around the body and connected to a common manifold. As with all pneumatic systems the valve switches the fluid supply from one side of the actuator to the other such as to reduce the error. The only other source of power is a d.c. electrical supply and the system is extremely simple. However Table 2.4-1 shows that for typical missile speeds the pressure recoverable from a pitot tube is low.

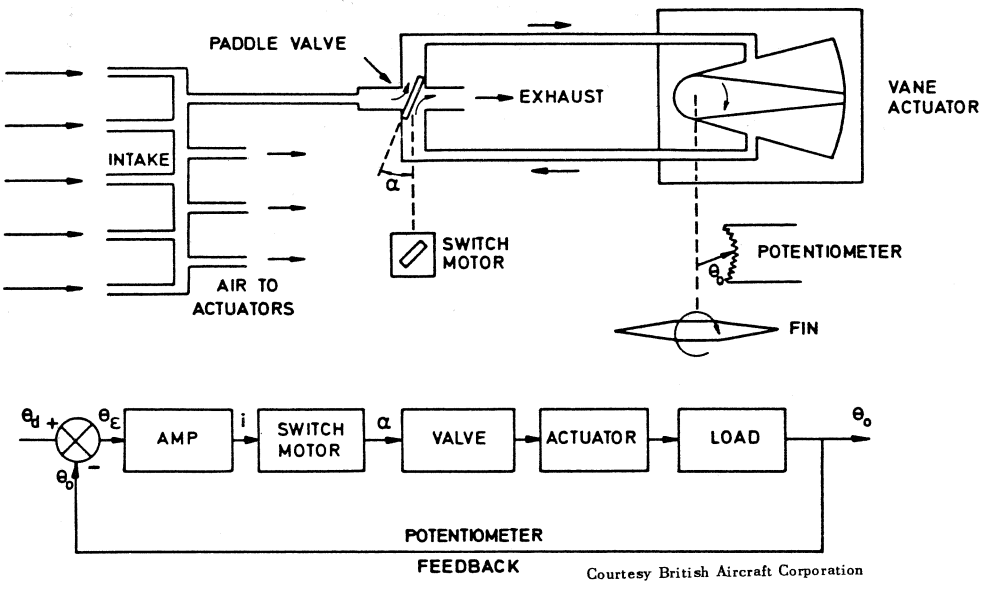


FIG 2.4-1 A ram air servo

TABLE 2.4-1 Stagnation pressure (MN/m^2) as a function of Mach Number and height

Height \ Mach Number	0.8	0.9	1.0	1.5	2.0	3.0
0 km	0.053	0.070	0.091	0.243	0.472	1.12
10 km	0.018	0.024	0.306	0.082	0.159	0.376

A ram air servo is unthinkable for speeds below about $M = 1.5$. This relatively low supply pressure means that, for a given servo actuator torque, the actuator size will be larger; this has a beneficial side effect in increasing the pneumatic natural frequency.

The dynamic performance of a ram air system will vary with Mach number and altitude and this tends to follow the aerodynamic characteristics of the missile. This should not create any new autopilot design problems. Since the size of the complete servo system does not increase with flight time (except for the d.c. electric supply) as does that of a conventional hot or cold pneumatic system which requires a larger gas generator or gas bottle, it is found that a ram air system can offer slight weight and size advantages for long times of flight. However, there still remains the problem of providing an alternative power source (say a small cold gas bottle) for control during the boost stage.

2.5 HYDRAULIC SERVOS

Conventional valve-controlled hydraulic actuators have been used in many systems. If a supply pressure of 150 MN/m^2 or more is used the actuator size is very small indeed and torque to inertia ratios are exceptionally high. Hydraulic natural frequencies are higher on account of the much greater bulk modulus of hydraulic fluid compared with that of say nitrogen at 50 MN/m^2. Very high performance servos can be designed using hydraulic power; bandwidths of 50 Hz or more can be obtained. They are an obvious choice when large inertias are involved, such as moving wings. Also, if a recirculating system as described by Walters (4) is used a hydraulic system will probably be the lightest of all for a long time of flight. A non-circulating system where oil is expelled from a tank by an explosive charge has been used in Seacat which has a short time of flight and moving wings; used oil is discharged to atmosphere.

In a recirculating system high pressure oil can be obtained by means of a cool cordite gas generator which drives a turbine coupled to a gear pump (Thunderbird). Alternatively a mono-fuel such as isopropyl nitrate can be burnt to provide gas to operate a two cylinder reciprocating oil pump. None of these items are cheap and when one adds the cost of the filters, actuators, relief valves, and control valves (the most expensive item) it will be found that a small hydraulic servo system for a tactical missile is a very expensive item indeed. Hydraulic servos are not made commercially below about 8 kW and tactical missile servos rarely require such large powers. It is unlikely that a hydraulic servo would be considered for a missile unless the missile weight is greater than about 400 kg and the time of flight more than a minute.

2.6 ELECTRIC SERVOS WITH D.C. MOTORS

A conventional permanent magnet d.c. servo motor apparently compares very unfavourably with a pneumatic or hydraulic motor on a power/weight basis. However, the power output quoted by the manufacturer of an electric motor is the *rated* power, i.e. based on the assumption of continuous operation resulting in a given temperature rise. For a flight time of a minute or less an electric motor can be considerably over-rated. Most conventional permanent magnet motors can accept an armature current three or four times the rated current; currents larger than this result in partial demagnetisation. Some missiles using d.c. electric motors use motors with wound fields control being applied to the armature. This results in a slightly heavier motor and

a small loss in efficiency but very large currents can be accepted for short periods with no permanent weakening of the field. Alternatively motors with samarium cobalt permanent magnets can produce a short term output of nearly 0.4 kW/kg and a zero speed maximum acceleration of nearly 1.5×10^5 rad/sec^2; this is considerably higher than that obtainable from basket-wound or printed-circuit low-inertia armature motors. The cost of the permanent magnets means that these motors are expensive. Low-inertia ironless rotor motors however have very short overload ratings as the absence of iron results in a rapid rise in the temperature of the windings when on overload. Such motors could be attractive for missiles with flight times of less than about 20 seconds.

For a given motor torque the maximum load acceleration is obtained when $G = \sqrt{J_\ell/J_m}$, G being the gear ratio. The equivalent inertia at the motor shaft is then $J_m + J_\ell/G^2$ and for a motor torque T_m the acceleration of the motor is $T_m/(J_m + J_\ell/G^2) = T_m/2\,J_m$. The acceleration of the load $\ddot{\theta}_\ell$ is therefore given by

$$\ddot{\theta}_\ell = T_m/2J_mG = T_m/2\,\sqrt{J_mJ_\ell}$$

and since J_ℓ is a fixed quantity it follows that

$$\ddot{\theta}_\ell \propto T_m/\sqrt{J_m}$$

It follows therefore that the greatest load accelerations are possible with motors having the best torque^2/inertia ratio.

2.7 OTHER ELECTRIC SERVOS

The power developed by a servo motor in controlling a missile control surface is consumed mainly in accelerating and decelerating the combined inertias of the motor and load; the power necessary to brake the load has to be supplied and is then dissipated as heat in the motor. A regenerative drive would be far too expensive for such a simple expendable device. A method of reducing power requirements and eliminating the effect of the motor inertia is to run a motor (preferably with increased inertia!) and control the load by clutches or by a clutch/capstan drive. When the load is accelerating power is absorbed from the motor and when the load is decelerating the power flow is reversed; the torque is reversed but the motor continues rotating in the same sense. Over a period of time the motor is required to make up the friction losses and the losses in the clutches when they are slipping. A very compact servo (5) is shown on outline in Fig 2.7-1. A double speed reduction is obtained by

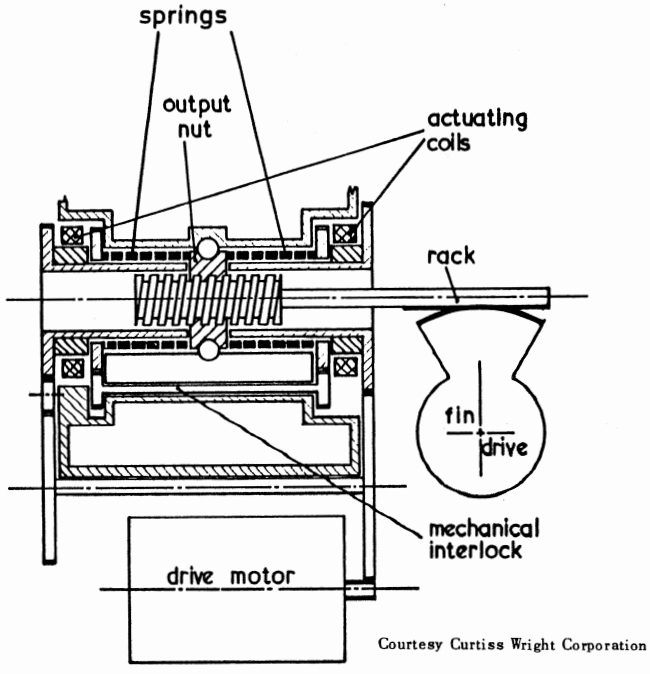


FIG 2.7-1 A clutch/capstan fin actuator

spur gears and a nut and screw drive. Two contra-rotating drums are in constant motion from a direct drive from the motor. A spring is wrapped round each drum and is attached to the nut at one end and one member of a clutch at the other. Actuation of one clutch provides positive torque to the final drive, the capstan acting as the main power amplifier; actuation of the other clutch provides torque in the opposite sense. A mechanical interlock prevents the simultaneous operation of both clutches. When the error is zero the springs act as brakes locking the output shaft. In this design the maximum power output of the motor was reduced to less than 60% of the maximum power demand. Alternatively, the space available for the servo components may suggest that the capstans be dispensed with and the main power amplification be provided by magnetic particle clutches alone; the heat dissipation in the clutches may now be a limiting factor. However, the fact that one cheap motor can be shared between two or four servos is a considerable advantage.

How do solenoids for direct actuation of control surfaces compare with other forms of electric actuation? Surely for sheer simplicity, cheapness and reliability there is no equal in the field of electro-mechanical energy

conversion. However, on a power/weight basis they compare unfavourably with d.c. motors. The fundamental reason why there is a difference is that although both devices utilise the force generated between two surfaces due to electric currents and electro-magnetic fields we use these forces over and over again in a rotary machine thus enabling large powers to be converted. In a solenoid the force is a large attractive one which is used once only in a working stroke; the force also varies considerably with the changing air gap. However, for small missiles and very short times of flight the simplicity and compactness of direct solenoid operation will certainly be attractive. The use of solenoids in homing head actuation is remarked upon in Chapter 8. And finally we note that although all missiles will need some form of electric power supply, those using electric actuation will need much more. It is difficult to provide the perfect power supply for a missile with a maximum flight time of 20 or even 10 seconds as ideally we would like a simple cheap battery which provides a near-constant voltage and which is capable of giving up the whole of its stored energy in this very short time. Thermal batteries contain an electrolyte which at ordinary temperatures is an inert solid. When power is required a chemical heat source within the battery is ignited and this melts the electrolyte. Activation times are usually less than a second. They have excellent shelf life and reliability but suffer from considerable voltage droop as discharge continues. If the flight time is less than 40 seconds or so some of the battery capacity has to be wasted. There appears to be no hope at the present of providing a thermal battery which is efficient for extremely short discharge times, although of course they are widely used in systems with maximum flight times of say 12 seconds. Such systems do not usually use electric actuation. The alternative to the thermal battery is the silver-zinc battery where the potassium hydroxide electrolyte is forced into the cells just before use. These batteries are best used when the discharge time is several minutes.

2.8 SOME TENTATIVE CONCLUSIONS

Missile servos are rarely cheap, especially for the high performance surface-to-air and air-to-air missiles which require such a large bandwidth and high maximum fin rate. The author is of the opinion that it is not possible to meet the specification for servos in this class with electric actuation of any sort. Very high performance servos can be designed using hot or cold gas and for the heavier medium-range missiles hydraulic servos usually provide

the lightest and most compact solution - and the most expensive. Systems designed to hit stationary or slow moving targets require relatively low bandwidth autopilots and if the noise levels are not expected to be high a modest servo specification will be sufficient enabling electric actuation to be used. One method of reducing servo power requirements is to use twist and steer, canard controls and a head containing the servos and control fins decoupled from the main body in roll. Ailerons will then have a very small inertia to control and elevators need to operate in one sense only. Direct solenoid actuation has been used in such a system.

REFERENCES

1. POOLE Harmon A. Jr A stored gas actuator unit for the TOW missile. Proceedings of fluid power systems and component conference. S.A.E. 1965.

2. BURROWS C.R. Fluid power servomechanisms. Van Nostrand Reinhold Coy 1972.

3. SHOESMITH T. Development of a high speed solenoid and drive amplifier. Royal Aircraft Establishment Technical Report No 69206.

4. WALTERS R. Hydraulic and electro-hydraulic servo systems. Iliffe Books Ltd 1967.

5. HALL J.L. and SCHREIBER C.R. Flight control actuators for missile applications. American Society of Engineers (N.Y.) 1968 Conference in aviation and space-progress and prospects.

6. ABATE S.J. The future of electrical acutation of missile control surfaces. M.Sc. thesis, Guided Weapon Systems Course, Royal Military College of Science, Shrivenham. 1973.

CHAPTER 3

MISSILE CONTROL METHODS

3.1 INTRODUCTION

Before going into mathematical detail concerning the motion of a missile in space as a result of guidance commands, some definitions and discussion are desirable. For instance, are we going to let the missile roll freely or are we going to control its orientation in roll? Are we going to manoeuvre the missile by moving control surfaces or are we going to alter the direction of thrust? This chapter is therefore mainly descriptive and discusses ways and means.

It is convenient to start with a definition of the task of a missile control system. It is one of the tasks of the guidance system to detect whether the missile is flying too high or too low, or too much to the left or right. It measures these deviations or errors and sends signals to the control system to reduce these errors to zero. The task of the control system therefore is to manoeuvre the missile quickly and efficiently as a result of these signals. Suppose the guidance "sees" the missile at m relative to its own boresight and that we interpret this to mean that the missile is too far to the right and too low. In a cartesian system the guidance angular error detector produces two signals, a left-right signal and an up-down signal which are transmitted to the missile by a wire or radio link to two separate servos, say rudder servos and elevator servos. Fig 3.1-1 shows that this same information could be expressed in polar co-ordinates i.e. R and ϕ. If the same information is expressed in another way then the control system must

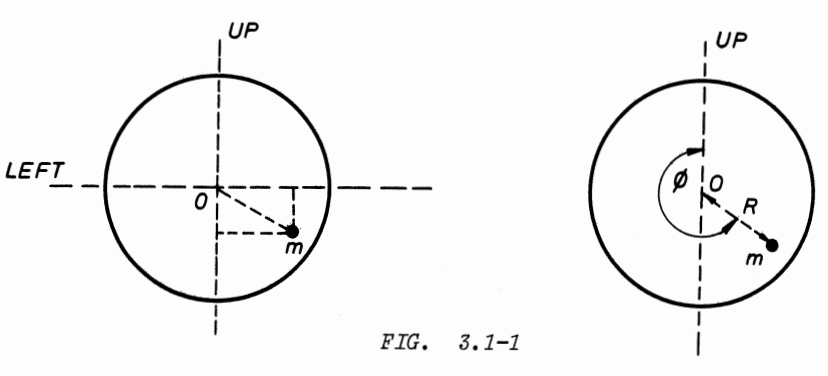


FIG. 3.1-1

be mechanised differently. The usual method is to regard the ϕ signal as a command to roll through an angle ϕ measured from the vertical and then to manoeuvre outwards by means of the missile's elevators. The method of control compatible with polar commands is called twist and steer.
And finally, by way of introduction, we should perhaps be rather more specific concerning the words "elevators" and "rudders"; and "ailerons" should also be defined. Guided missiles usually have one or two axes of symmetry. The reasons are discussed in Chapter 4. If a missile has four control surfaces as shown in Fig 3.1-2 one regards surfaces 1 and 3 as elevators and 2 and 4 as rudders even if the missile should roll subsequently.

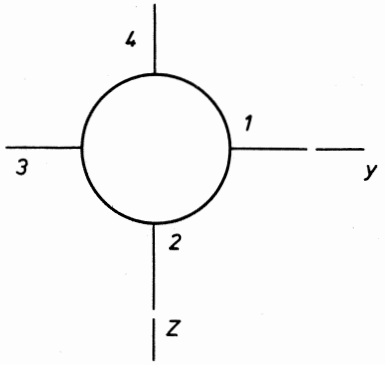


FIG. 3.1-2 Control surfaces looking from rear of missile

If 1 and 3 are mechanically linked together such that a servo must impart the same rotation to both then these surfaces are elevators pure and simple. The same argument applies to the rudders. Suppose now surfaces 1 and 3 each have their own servo it is possible for them to act as ailerons. If looking in the direction y one surface is rotated ξ^{o} clockwise and the other surface ξ^{o} anti-clockwise then a pure couple is imparted to the missile about the fore and aft axis and this will tend to make the missile roll. Such control surfaces are now called ailerons. We can double the power of the ailerons by doing the same thing to control surfaces 2 and 4. If now the aerodynamics are linear i.e. the normal forces are proportional to incidence then the principle of superposition applies. Commands for elevator, rudder and aileron movements can be added electrically resulting in unequal movements to opposite control surfaces. In this way we have the means to control roll motion as well as the up-down (i.e. pitch) motion and left-right (i.e. yaw)

motion. Alternative methods of control are shown in Fig 3.1-3.

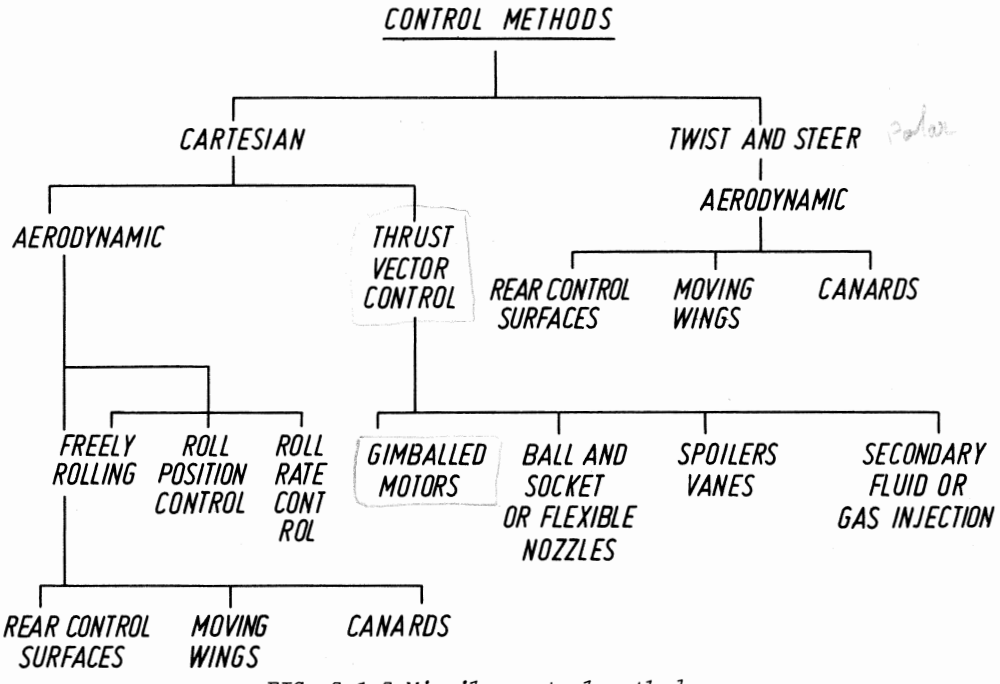


FIG 3.1-3 Missile control methods

3.2 ROLL CONTROL

With cartesian control ideally one would like the missile to remain in the same roll orientation as at launch during the whole flight. Up-down guidance signals if sent to the elevator servos should then result in a vertical manoeuvre of the missile; and left-right signals if sent to the rudder servos should result in a horizontal manoeuvre. However a missile is not designed like an aeroplane and there is no tendency to remain in the same roll orientation. In fact it will tend to roll due to any of the following causes:

(a) accidental rigging errors which cannot be eliminated entirely.

(b) asymmetrical loading of the lifting and control surfaces in supersonic flight which occur when pitch and yaw incidences occur simultaneously and are not equal. This effect can be considerable but can be minimised by good design and tends to be small for long thin missiles. The roll moment from rear stabilising surfaces when asymetric shock waves impinge on them can be more or less eliminated by mounting these surfaces on a collar which is free to rotate about the missile fore and aft axis.

(c) atmospheric disturbances especially if the missile is flying close to the ground.

If the guidance error detector is on the ground up-down and left-right signals can be implemented correctly if a roll gyro and resolver exist in the missile to ensure that the commands are mixed in the correct proportions to the elevators and rudders. However, it will be demonstrated in a later chapter that high roll rates cause cross-coupling between the two channels and tend to unstabilise the system. Conversely, if the guidance error detector is in the missile (e.g. all homing systems), then the guidance system and the control system share the same reference axes; if the guidance error detector rotates so does the control system and the necessity of resolving guidance signals due to missile rotation does not arise. However, there are many occasions when the system designer would wish to roll position control (i.e. stabilise) the missile:

(a) If the missile is likely to be guided by radar at a low angle over the ground or sea, the use of vertically polarised guidance commands together with vertically polarised aerials in the missile will assist the missile receiver in discriminating against ground or sea reflections.

(b) If the missile is a sea skimmer controlled in height by a radio altimeter, it will be necessary for the altimeter to point downwards $\pm 15^{o}$ or so. However, if sufficient apertures are available it may be possible to install an all-round looking aerial system.

(c) If the missile is a homer and there is a mechanical drive from the missile to the homing head then the target may be lost if there is a high roll rate and the homing head is looking sideways; the inertia of the drive motor and friction in the drive will tend to make the head follow the missile motion.

(d) Due to the aerodynamic shape of the missile it may be considered inadvisable to risk the possibility of high roll rates due to the adverse affect on system accuracy already mentioned. If high roll rates are the danger then the missile can be roll-rate stabilised instead of roll-position stabilised.

It should be noted that a twist and steer form of control involves roll position stabilisation; or more specifically it involves roll position demand. The details of implementing roll control are discussed under autopilot design.

3.3 AERODYNAMIC LATERAL CONTROL

With a cartesian control system the pitch control system is made identical to the yaw control system so we need discuss one channel only; in this respect the nomenclature differs from that used in aircraft. With missiles lateral movement usually means up-down or left-right. With polar control one rolls and elevates. The following remarks apply to the elevation channel in twist and steer missiles also.

Rear control surfaces

The majority of tactical missiles have fixed main lifting surfaces (often called wings) with their centre of pressure somewhere near the missile centre of gravity and rear control surfaces. With subsonic missiles it may be more efficient to use the controls as flaps immediately behind the wings as the flap controls the circulation over the whole surface. With supersonic flow the control surface cannot affect the flow ahead of itself and therefore it is placed as far to the rear as possible in order to exert the maximum moment on the missile. Rear control surfaces often make for a convenient arrangement of components. Usually it is desirable to have the propulsion system placed centrally in the missile so that centre of gravity shifts due to propellant usage are minimised. It is convenient and sometimes essential to have the warhead and fuze at the front together with any associated electronics including the guidance receiver. This leaves the control system to occupy the rear end with the propulsion blast pipe passing through its centre. If there are four servos it is not difficult to design a neat servo package round this pipe.

When considering the lateral forces and moments on missiles it is convenient first of all to consider the combined normal forces due to incidence on the body, wings and control surfaces as acting through a point on the body called the centre of pressure (c.p.) and to regard the control surfaces as permanently locked in the central position. If the c.p. is ahead of the centre of gravity (c.g.) then the missile is said to be statically unstable. If it coincides with the c.g. then it is said to be neutrally stable and if it is behind the c.g. it is said to be statically stable. This of course is the reason why feathers are placed at the rear end of an arrow to move the c.p. aft. These three possible conditions are shown in Fig 3.3-1 to 3.3-3. The missiles are shown with a small incidence i.e. the body is not pointing in the same direction as the missile velocity vector U_m. In the unstable condition any perturbation of the body away from the direction of the velocity

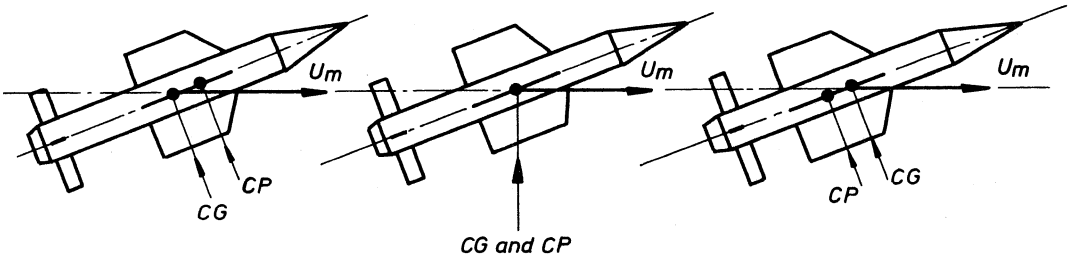


FIG. 3.3-1 Unstable *FIG. 3.3-2 Neutrally stable* *FIG. 3.3-3 Stable*

vector results in a moment about the c.g. which tends to increase this perturbation. Conversely in the stable condition any perturbation of the body direction results in a moment which tends to oppose or decrease this perturbation. The distance of the c.p. to the c.g. is called the static margin. Since lateral force and hence lateral manoeuvre by aerodynamic means is obtained by exerting a moment on the body such that some incidence results it follows that if the static margin is excessive, the missile is unnecessarily stable and control moments will be relatively ineffective in producing a sizeable manoeuvre. There has to be a compromise here between stability and manoeuvrability. Now consider a missile whose forward speed is constant, with a steady body and wing incidence of β and a control surface movement from the central position of ζ. We will consider motion in the horizontal plane only and assume the missile is not rolling; the effects of gravity are zero on this plane. Fig 3.3-4 shows the normal force N due to the body, wings and rear control surfaces assumed to be in the central position; this force N acts through the c.p. But there will be an additional force N_c due to the control surfaces being deflected by an amount ζ. Neglecting the small damping moment due to the fact that the missile is executing a steady turn, this picture can represent dynamic equilibrium if the rudder moment $N_c \ell_c$ is numerically equal to Nx^* where x^* is the static margin. If $\ell_c/x^* = 10$ say then $N = 10\ N_c$, and the total lateral force = $9\ N_c$. Note that this force is in the opposite sense to N_c. Since x^* is typically 5% or less of the body length it is easily seen that a small absolute change in the static margin can affect the manoeuvrability of the missile. Thus the standard method of obtaining a large lateral force on a missile is to have a large moment arm by placing the control surfaces as far from the c.g. as possible. The reader may wish to verify that if the c.p. of the wings and body alone is at the c.g. then ζ^0 of control surface movement will produce the same amount of body incidence. If the c.p. as just defined is in front of the c.g. then

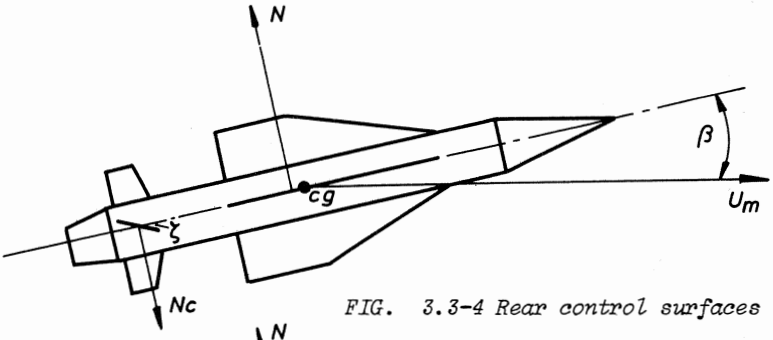


FIG. 3.3-4 Rear control surfaces supersonic

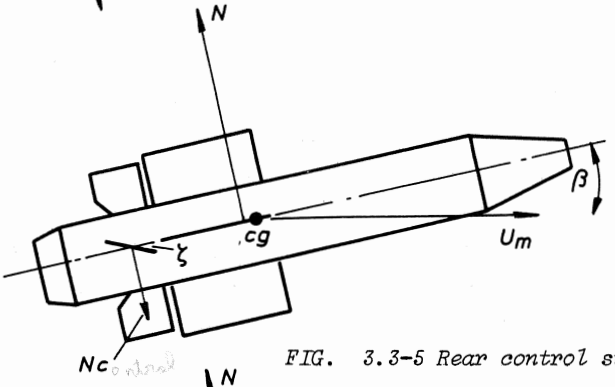


FIG. 3.3-5 Rear control surfaces subsonic

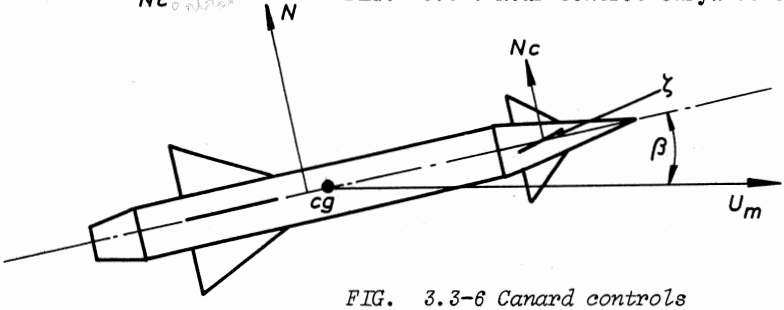


FIG. 3.3-6 Canard controls

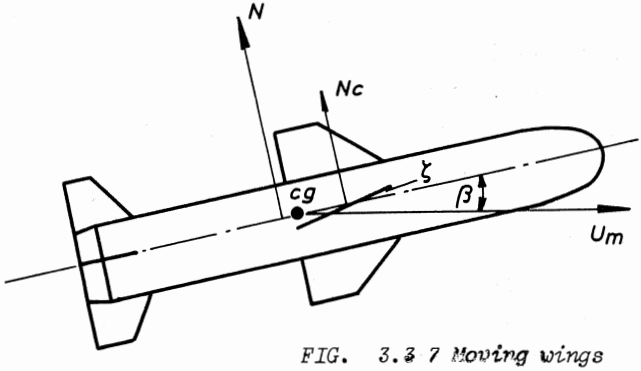


FIG. 3.3 7 Moving wings

ζ^0 of rudder will induce more than ζ^0 of body incidence. If this c.p. is behind the c.g. then less than ζ^0 of body incidence will result. If a missile has no autopilot (i.e. no instrument feedback) a considerable static margin has to be allowed to ensure stability in flight, say 5% or more of the overall length. With instrument feedback zero or even negative static margins can be used, thus assisting manoeuvrability. Before leaving rear controls it should be noted that the overall c.p. can never be regarded as in a fixed position. The c.p. of the body in particular will vary with incidence and with Mach number.

Forward control surfaces

Since the main object of siting a control surface is to place it as far from the c.g. as possible, a position as far forward as is practicable appears a logical choice. Forward control surfaces are often called "canards" named after ducks who apparently steer themselves by moving their head. Fig 3.3-6 shows another possible case of dynamic equilibrium. In this case it is seen that the lateral force due to the missile as a whole now adds to the force due to the deflection of the control surface and therefore if $\ell_c/x^* = 10$ as before, then the total normal force is 11 N_c compared with 9 N_c with rear controls. Also, the final sense of the total normal force is in the same sense as the control force. Canards therefore are slightly more efficient in the use of lateral control forces. If the reader thinks that canards will automatically render the missile unstable he will note that the canard controlled missile has been drawn with the main lifting surfaces rather further aft to make the overall c.p. aft of the c.g. It is the position of the overall c.p. relative to the c.g. which is the stability criterion. Since canards appear to be superior to rear controls why do we find so many missiles with rear controls? Firstly, we shall see that with feedback instruments in a control system the static margin can be made zero or even negative whilst maintaining overall stability; so the difference in total normal force available can often be almost negligible. Secondly, there is the question of convenience in packaging as already mentioned which usually favours rear controls. Finally, and this is probably the chief reason, the downwash from canards on to the main lifting surfaces can, in many configurations, nullify any attempt to control the missile in roll. A long thin missile suffers less in this respect than a short one. There are two ways around this problem. If the missile is a twist and steer one the small front portion of the missile can be attached to the main body by means of a

bearing which allows the rear body to rotate freely thus uncoupling the head from rolling moments induced in the wings. An alternative is to mount the wings on a collar which allows them to rotate freely around the body as already mentioned.

Moving wings

To use servos to move the main lifting surfaces and employ small fixed rear stabilising surfaces is unusual. There could be the rare occasion when the servos are most conveniently placed near the centre of the missile. However the main reason for adopting this configuration would be, for a given lateral acceleration, to minimise the body incidence. If the propulsion system is a ram jet the air intake is likely to choke if the body incidence is large, say 15^{o} or more. Alternatively, if the missile is a sea skimmer with height control governed by a radio altimeter it may be necessary to specify a maximum body incidence. In Fig 3.3-7 the normal force from the wings for a given incidence is shown as half that due to wings, body and stabilising surfaces combined, a situation which is possible for fairly large fixed stabilising surfaces. If the distance of the c.p. of the wings forward of the c.g. is twice the static margin then the steady state body incidence is equal to the wing deflection, giving a total normal force three times the original normal force from the wings. A maximum total wing incidence of say 20^{o} means that the maximum body incidence is only 10^{o}. However there are some distinct penalties involved in the use of moving wings. Clearly the servos will be appreciably larger to cope with the increased inertia of the load and the larger aerodynamic hinge moments. Also, it has already been demonstrated that moving wings are an inefficient way of producing a large normal force due to the small moment arm available. Owing to the fact that the whole bending moment at the wing root has to be taken by a shaft, the wing will have to be designed much thicker around the mid chord. This not only increases the structure weight but at supersonic speeds it will increase the drag, the pressure drag varying with the thickness-to-chord ratio squared. It is desirable to make the centre section of the missile body square in cross section to eliminate a large wing-body gap when the wing is deflected; such a gap considerably reduces the generated normal force. And finally since the moment arm is small, the position of the c.g. is critical as a small shift will make an appreciable change in the control moment arm. Nevertheless, if the maximum g requirements are low and the speed is subsonic, such as for an anti-ship missile, the overall weight penalty may not be excessive if small moving wings are used.

3.4 AERODYNAMIC POLAR CONTROL VERSUS CARTESIAN CONTROL

The great majority of tactical missiles are required to have the same degree of manoeuvrability vertically and horizontally; this is why the conventional configuration is a cruciform one with two equal pairs of lifting surfaces and two pairs of control surfaces. If no form of roll control is required only two servos are necessary provided opposite control surfaces are linked together. If some form of roll control is required at least three servos are necessary and in practice four identical servos are usual. If only three were used and they were all the same size the two controlling only one fin each would be overpowered. The advantage of adopting a twist and steer method of control is to use only one pair of lifting surfaces and one pair of control surfaces. This cuts down weight and drag and such a configuration will certainly be advantageous for horizontal storage between decks of a ship or for launching under the wings of an aircraft. As an example of the drag bonus, about a half of the total drag of a conventional supersonic missile comes from the four wings and four control surfaces. Clearly there are advantages. The method of manoeuvring the missile is this. The ϕ command goes as a positive command to one control surface and a negative command to the other. This causes the missile to roll. The R command goes to both surfaces always as a positive demand. The intention is to make the response in roll so fast that the commands can be applied simultaneously which makes for simplicity. Nevertheless, manoeuvre cannot be so efficient and fast as with cartesian control. There is also a possibility that in certain situations the performance of a twist and steer system could be distinctly inferior to a cartesian one. Suppose in a surface to air system a target is approaching nearly head on. The missile has to pull very little g but if there is much noise in the system the missile could easily be rolling backwards and forwards through 180^{o} or so because the target appears to be at 12 oclock and then 6 oclock. The effect on system accuracy is difficult to predict under such circumstances. Bloodhound is the well known example of polar control and the arguments for its adoption stem primarily from the choice of a ram jet for the propulsion motor. Now if two ram jets external to the main body are used there is room for only one pair of wings. At the time the air intakes could accept only a small body incidence due to interference from the body so moving wings were adopted (1). These two wings are therefore used as ailerons and elevators at the same time and are controlled by separate servos; the word "elevons" is sometimes applied to such control surfaces.

Best (2) has demonstrated some of the difficulties in ensuring a stable system with polar control used in a homing system, and in assessing the system performance. With cartesian control it is possible to resolve target and missile motion into two planes and to consider the pitch and yaw channels as independent two-dimensional problems. This simplification is not possible in the case of polar control; indeed the equations of motion which result are, in general, not susceptible to analysis. Detailed three-dimensional simulation has to be employed. Bearing in mind the fact that cartesian control must be a quicker method of moving laterally in any one direction and that analysis of the performance of cartesian systems is simpler it is understandable why the great majority of missile systems use this method. A later missile, Seadart, uses an integral ram jet with the air intake in the nose. Since there is no interference from the main body larger incidences can be accepted with such an arrangement. Moving wings are not necessary to reduce body incidence and therefore a conventional cruciform missile using cartesian control and rear control surfaces was adopted; it was considered that an integral ram jet was not feasible with Bloodhound.

3.5 THRUST VECTOR CONTROL

A completely different method of steering a missile is to alter the direction of the efflux from the propulsion motor and such a method is known as thrust vector control (TVC). This method of control is clearly not primarily dependant on the dynamic pressure of the atmosphere, but on the other hand it is inoperative after motor burn-out. It will be argued later that in many situations there are advantages in having a boost-coast velocity profile. TVC is therefore likely to have a limited application. The following situations make TVC essential or desirable:

(a) It is essential to use TVC in the vertical launch phase of all intercontinental ballistic missiles as these missiles whose total weight is well over 90% fuel have to be launched extremely gradually to avoid dynamic loading. Aerodynamic controls would be completely ineffective for some time and the missile would topple over due to a small inevitable thrust misalignment unless an attitude sensor and TVC were used.

(b) If a missile is separated some distance from its controller such as in the anti-tank system Swingfire and rapid gathering is required to achieve a short minimum range then it must be possible to manoeuvre the missile almost immediately after launch.

(c) In a short range air-to-air missile, one may be trying to hit a fast crossing target with no aim-off and with a flight time of a few seconds. The exceptional manoeuvrability one can obtain with TVC would give the system a better coverage.

(d) It can be argued that some systems would be cheaper and simpler if one launched vertically and then turned over rapidly, thus eliminating an expensive and heavy launcher.

(e) Vertical launch followed by a rapid turnover is an attractive concept for missiles carried and launched from a vehicle; 360^{o} arc of fire is obtainable, and storage and reloading is almost certainly facilitated.

(f) Submarine launched missiles surfacing in different sea conditions may well need very early course correction.

In order to appreciate the degree of manoeuvrability that can be achieved let us consider the requirement to turn the body of a missile through say 90^{o} as soon as possible after launch i.e. when the missile speed is low. Aerodynamic forces will be small during this period and will be neglected in the following calculation. Fig 3.5-1 shows the forces acting on the missile if the thrust T is deflected through an angle γ.

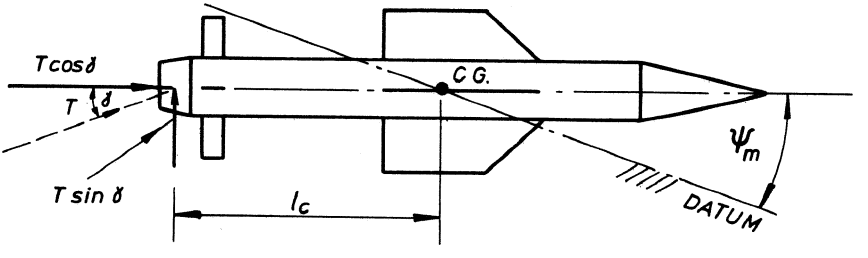


FIG 3.5-1 Forces due to thrust vectoring

If for example the mass of the missile is 30 kg, the moment of inertia about the c.g. is 1 kg m^2 (both assumed constant), the thrust is 1500 N and the moment arm ℓ_c is 0.35 m then for $\gamma = 4^{o}$

$$\ddot{\psi}_m = - T \ell_c \sin \gamma / J \quad \leftarrow$$

$$= -1500 \times 0.35 \times 0.07 = -36.8 \text{ rad/sec}^2$$

Assuming it is necessary to bring the body to rest at the end of the turn the perfect control system will accelerate the body through 45^{o} at maximum acceleration and then apply full braking torque for the same time. On these slightly optimistic assumptions the time for a complete turn from rest to rest will be found to be about 0.4 seconds; and this is for a comparatively small

thrust deflection. It should be noted that although the body has been turned through 90^{o} the velocity vector will have been turned through a smaller angle. To turn the velocity vector through a given angle requires initially a body turn greater than this in order to destroy any momentum already gained in the original direction. The difference between a horizontal turn and one in the vertical plane is that in the latter case gravity helps to destroy the vertical component of momentum and therefore a tight turn is slightly easier in the vertical plane.

There is, as with aerodynamic control, additional complication if roll control is required. Two ball and socket nozzles can produce a roll moment if they are moved differentially in one plane. Pitch and yaw control is obtained by moving them together in the desired plane each controlled by two servos. Three servos can be used if one moves both nozzles up-down for pitch and two independent servos move them left-right for yaw and roll control.

3.6 METHODS OF THRUST VECTORING

As indicated in Fig 3.1-3 there are several methods of directing the thrust of a rocket motor and each method has advantages and disadvantages which may or may not recommend it for a particular application.

Gimballed motors (Fig 3.6-1)

When liquid propellants are used the fuel and oxidiser are fed under pressure to one or more combustion chambers and expanded to atmosphere through convergent-divergent nozzles. If the combustion chambers are mounted in gimbals and controlled in position by servos the direction of thrust can be controlled. If a pair are moved differentially some control in roll is also obtained. The method is essentially loss-free and the loss of axial thrust is merely a cosine effect which is inevitable. As yet largely due to their toxidity liquid propellants are rarely used for tactical missiles despite their high specific impulse.

Moving nozzles (Figs 3.6-2 and 3.6-3)

If a solid propellant is used the next best thing to a gimballed motor is to preserve the smooth internal lines of a convergent-divergent by using a flexible nozzle or a ball and socket joint. In the former one attaches the venturi to the motor case by means of a flexible rubber mounting which is very stiff axially but relatively compliant in the pitch and yaw planes. 1^{o} of nozzle movement produces 1^{o} of thrust deflection. This is a good method for thrust deflections of 4^{o}-5^{o} but large torques are required to move a nozzle

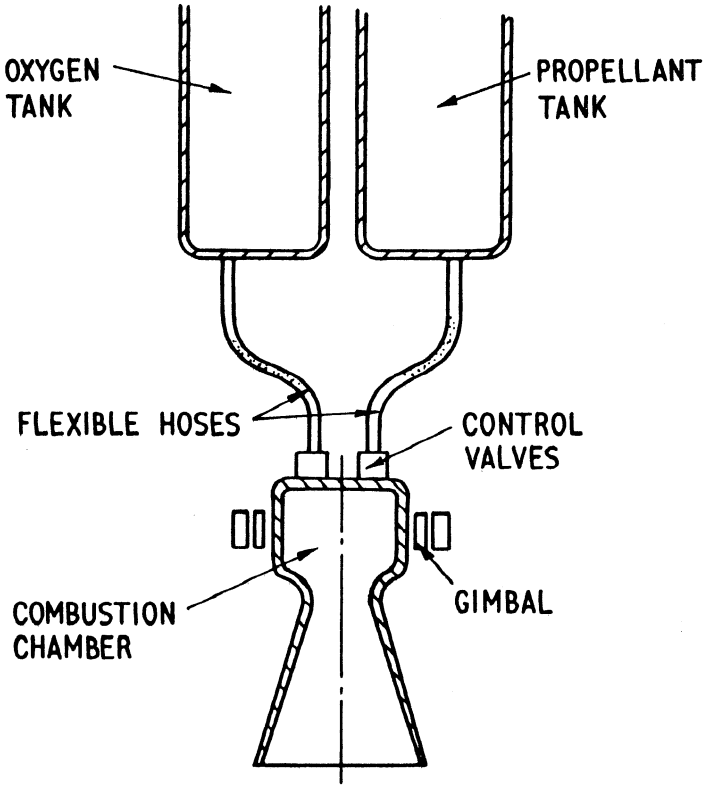


FIG 3.6-1 Gimballed motor

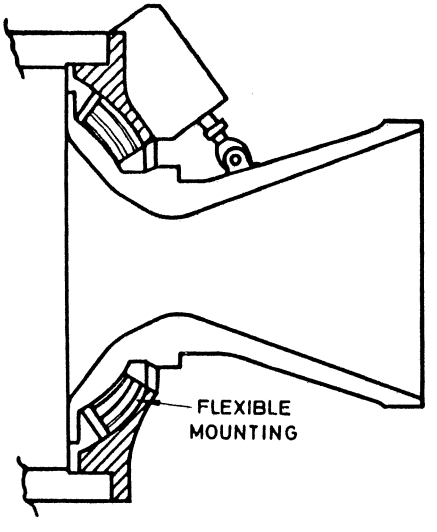


FIG 3.6-2 Flexible nozzle

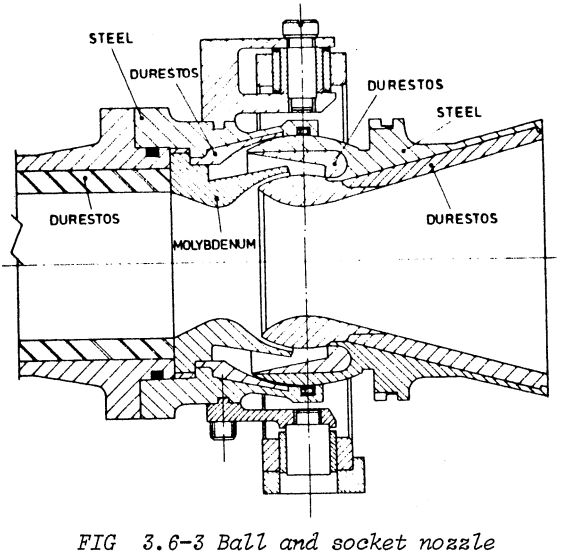


FIG 3.6-3 Ball and socket nozzle

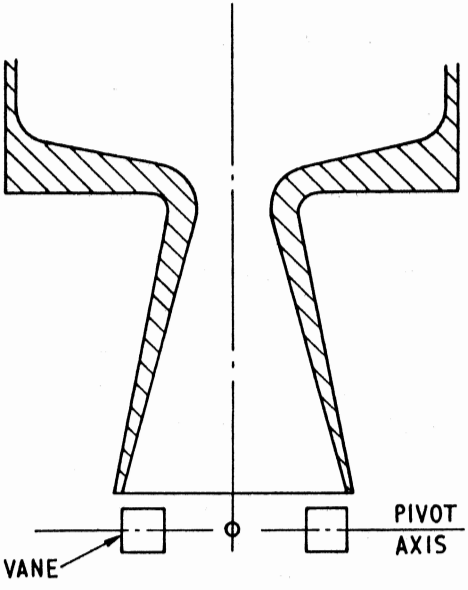


FIG 3.6-4 Moving vanes

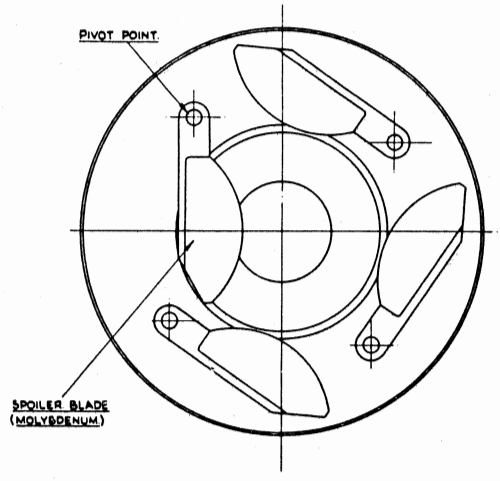


FIG 3.6-5 Semaphore spoilers

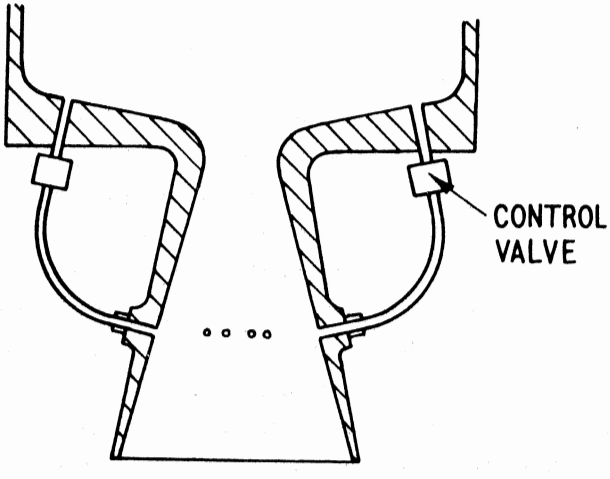


FIG 3.6-6 Gas injection

much more than this.

An obvious method is to use a ball and socket joint with some form of low-friction seal. $\pm 20^{o}$ of nozzle movement can be obtained and like the flexible nozzle is suitable for pressures of up to 50 bar.

Interference methods (Figs 3.6-4 and 3.6-5)

There are many methods of altering the momentum vector of a jet by inserting an obstruction in the stream. Since the flow is invariably supersonic shock waves are produced and this causes a rise in pressure and a loss in velocity. Interference methods are relatively crude and lossy, but have the advantage that very small moving parts are required and therefore the servos are small. Perhaps the oldest jet deflector is the moving vane as used in the German V2 in World War II. Many surface-to-surface missiles including the American Pershing have used vanes. There is a severe erosion problem of course due to the high gas temperature and velocity. Graphite has been used and more recently tungsten and molybdenum. With two vanes nearly 24^{o} vane angle is needed for 10^{o} thrust deflection. Because the vanes are permanently in the jet stream about 8% of the thrust is lost with four vanes in the null position increasing to 20% with 20^{o} incidence with two of the vanes.

Another common method is the semaphore spoiler which the French have used on most of their anti-tank missiles. Very small servos are required. $\pm 22^{o}$ thrust deflection has been obtained with British designs but the axial thrust loss is not less than 1% per degree deflection.

Injection methods (Fig 3.6-6)

The object of injecting a liquid or gas into the motor venturi is to obtain a sideways component of resultant thrust greater than venting radially outwards direct to atmosphere. Even if an augmentation of side-thrust is obtained the method fails if appreciable axial thrust losses result. Injection of a fluid like Freon is attractive as it has a high specific gravity, is inert and stores well as a liquid. This method is very convenient for small deflection angles of 2^{o}-3^{o} but for angles greater than this axial losses build up rapidly. Controlled injection of some of the propellant hot gas has the obvious advantage of simplicity and a reduction in the number of components. For reliability however the control of this very hot dirty gas would probably require fluidic amplifiers.

Summarising, interference methods (especially semaphore spoilers), despite the fact that they are not new are attractive because the servos are so small. They are strong contenders for short range missiles since the weight of

propellant carried is not large and an increase of say 10% in the propellant carried may not impose an overall weight penalty. They could also be considered for a longer range missile where the duty cycle is low. Moving nozzles are very efficient but the inertia and other forces to be overcome require considerably heavier servos. They are attractive for systems with longer times of flight. Injection methods, if they are considered at all could be used when very small deflection angles are adequate.

REFERENCES

1. FARRAR D.J. The Bloodhound J.R. Ac. Soc Jan 1959.
2. BEST D. Some problems of polar missile control J.R. Ac. Soc Aug 1960.
3. BURGESS B. Recent advances in thrust vector control for tactical missiles. The aeronautical journal Aug 1973.

CHAPTER 4

AERODYNAMIC DERIVATIVES AND AERODYNAMIC TRANSFER FUNCTIONS

4.1 NOTATION AND CONVENTIONS

The reference axis system standardised in the guided weapons industry is centred on the c.g. and fixed in the body as follows:

x axis, called the roll axis, forwards, along the axis of symmetry if one exists, but in any case in the plane of symmetry.

y axis, called the pitch axis, outwards and to the right if viewing the missile from behind.

z axis, called the yaw axis, downwards in the plane of symmetry to form a right handed orthogonal system with the other two.

Table 4.1-1 defines the forces and moments acting on the missile, the linear and angular velocities, and the moments of inertia; these quantities are shown in Fig. 4.1-1. The moments of inertia about O are defined as:

$$A = \Sigma\, \delta m\, (y^2 + z^2) \qquad (4.1\text{-}1)$$

$$B = \Sigma\, \delta m\, (z^2 + x^2) \qquad (4.1\text{-}2)$$

$$C = \Sigma\, \delta m\, (x^2 + y^2) \qquad (4.1\text{-}3)$$

The products of inertia are defined as:

$$D = \Sigma\, \delta m\, yz \qquad (4.1\text{-}4)$$

$$E = \Sigma\, \delta m\, xz \qquad (4.1\text{-}5)$$

$$F = \Sigma\, \delta m\, xy \qquad (4.1\text{-}6)$$

TABLE 4.1-1 NOTATION

	Roll axis x	Pitch axis y	Yaw axis z
Angular rates	p	q	r
Component of missile velocity along each axis	U	v	w
Component of force acting on missile along each axis	X	Y	Z
Moments acting on missile about each axis	L	M	N
Moments of inertia about each axis	A	B	C
Products of inertia	D	E	F

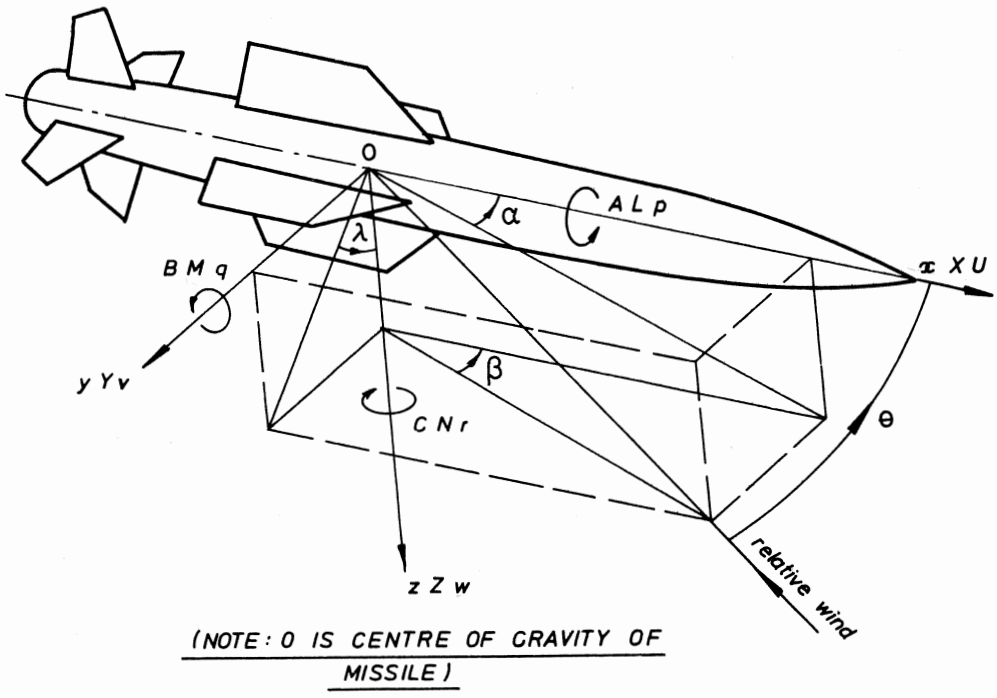


FIG. 4.1-1 Force, moment, etc conventions

The yaw plane is the Oy plane and the pitch plane is the Oz plane. The following angles are defined:

α: incidence in the pitch plane.

β: incidence in the yaw plane.

λ: incidence plane angle.

θ: total incidence, such that:

$$\tan \alpha = \tan \theta \cos \lambda$$

$$\text{and } \tan \beta = \tan \theta \sin \lambda$$

The reason why U, the missile velocity along the x axis, is denoted by a capital letter is to emphasise that it is a large positive quantity changing at most only a few per cent per second. The angular rates and components of velocity along the pitch and yaw axes however, tend to be much smaller quantities which can be positive or negative and can have much larger rates of change.

4.2 EULER'S EQUATIONS OF MOTION FOR A RIGID BODY

There are six equations of motion for a body with six degrees of freedom, three force equations and three moment equations. If the missile mass is m they are:

$$m\,(\dot{U} + qw - rv) = X \tag{4.2-1}$$

$$m\,(\dot{v} + rU - pw) = Y \tag{4.2-2}$$

$$m\,(\dot{w} - qU + pv) = Z \tag{4.2-3}$$

$$A\dot{p} - (B - C)\,qr + D(r^2 - q^2) - E(pq + \dot{r}) + F(rp - \dot{q}) = L \tag{4.2-4}$$

$$B\dot{q} - (C - A)\,rp + E(p^2 - r^2) - F(qr + \dot{p}) + D(pq - \dot{r}) = M \tag{4.2-5}$$

$$C\dot{r} - (A - B)\,pq + F(q^2 - p^2) - D(rp + \dot{q}) + E(qr - \dot{p}) = N \tag{4.2-6}$$

The first equation does not really concern us; we are interested in the acceleration perpendicular to the velocity vector as this will result in a change in the velocity direction. In any case in order to determine the change in the forward speed we need to know the magnitude of the propulsive and drag forces. Now consider equation (4.2-2). The term $-mpw$ is saying that there is a force in the y direction due to incidence in pitch ($\alpha = w/U$) and roll motion. In other words the pitching motion of the missile is coupled to the yawing motion on account of roll rate. The term mpv in equation (4.2-3) is also saying that yawing motion induces forces in the pitch plane if rolling motion is present. This is most undesirable as we require these two "channels" to be completely uncoupled. Ideally rudder movements should produce forces and moments in the yaw plane and result in yawing motion only; elevators should result in a manoeuvre in the pitch plane. Cross-coupling between the planes must contribute to system inaccuracy. To reduce these undesirable effects the designer tries to keep roll rates as low as possible, and in a simplified analysis one usually neglects the terms pv and pw if roll rates are expected to be small and incidence (v and w are proportional to incidence) is not large.

Now consider the moment equations. Ideally these should read

$$A\dot{p} = L;\ B\dot{q} = M;\ C\dot{r} = N$$

i.e. moments about a given axis produce angular accelerations about that axis. All other terms in these equations are cross-coupling terms and are undesirable from the point of view of system accuracy. We note that three out of four of the cross-coupling terms in each equation disappear if the products of inertia are zero. All three products of inertia are zero if there are two axes of symmetry, and two will be zero and the third will be

small if there is one axis of symmetry and the missile is reasonably symmetrical about another axis. With two planes of symmetry and a small roll rate therefore these equations reduce to:

$$m\,(\dot{U} + qw - rv) = X \tag{4.2-8}$$

$$m\,(\dot{v} + rU) = Y \tag{4.2-9}$$

$$m\,(\dot{w} - qU) = Z \tag{4.2-10}$$

$$A\dot{p} - (B - C)\,qr = L \tag{4.2-11}$$

$$B\dot{q} = M \tag{4.2-12}$$

$$C\dot{r} = N \tag{4.2-13}$$

The justification for neglecting the terms pq, pr, pv, pw is that the terms q, r, v and w are not large terms and if p is small then their products can be neglected. Equation (4.2-11) shows that there is zero coupling between the pitch and roll and yaw and roll motions if there are two axes of symmetry ($B = C$) and unless the missile is very unsymmetrical the cross-coupling should be weak.

4.3 CONTROL SURFACE CONVENTIONS

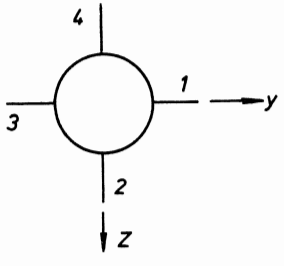


FIG. 4.3-1 Control surfaces as seen from the rear of the missile

Let the control surfaces be numbered as shown in Fig. 4.3-1. Deflections ξ_1, ξ_2, ξ_3, ξ_4 are defined as positive if clockwise looking outwards along the individual hinge axis. The following quantities are defined:

Aileron deflection $\xi = \frac{1}{4}\,(\xi_1 + \xi_2 + \xi_3 + \xi_4)$

or $\frac{1}{2}\,(\xi_1 + \xi_3)$ or $\frac{1}{2}\,(\xi_2 + \xi_4)$

if only two surfaces act differentially.

Elevator deflection $\eta = \frac{1}{2}\,(\xi_1 - \xi_3)$

Rudder deflection $\zeta = \frac{1}{2}\,(\xi_2 - \xi_4)$

The reader may wish to check that

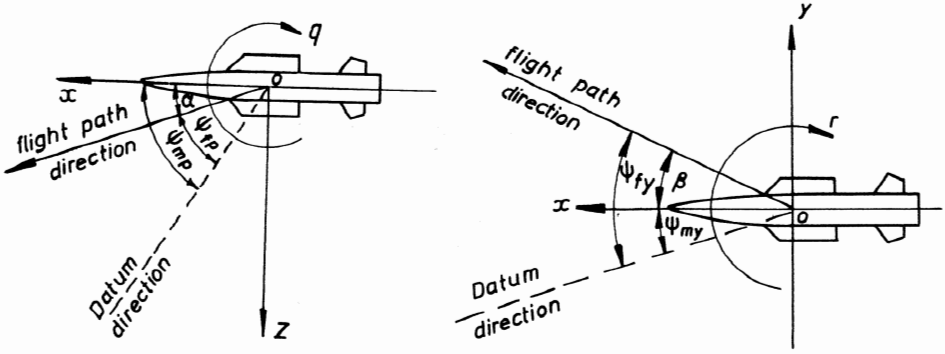
Positive aileron deflection produces an anti-clockwise moment about the x axis.

Positive elevator deflection produces a -ve force in the z direction and an anti-clockwise moment about the y axis.

Positive rudder deflection produces a +ve force in the y direction and a -ve moment about the z axis.

4.4 ANGLE EQUATIONS AND KINEMATICS

Since angular rates in pitch and yaw are +ve if viewed outwards along the y and z axes, it follows that angular positions must obey the same rule. Figure 4.4-1 shows an elevation of the missile so that angles in the pitch plane can be represented and a plan view to denote angles in the yaw plane.



PITCH PLANE YAW PLANE

FIG. 4.4-1 Angle conventions

Incidence in pitch $\alpha = \tan^{-1} w/U$ and for small angles $\alpha \simeq w/U$.

If the body direction is ψ_{mp} and the flight path direction ψ_{fp} then

$$\dot{\psi}_{mp} = q$$

$$\text{and } \psi_{fp} = \psi_{mp} - \alpha$$

$$\dot{\psi}_{fp} = \dot{\psi}_{mp} - \dot{\alpha}$$

$$\text{i.e. } \dot{\psi}_{fp} = q - \dot{w}/U$$

$$\text{and } U\dot{\psi}_{fp} = Uq - \dot{w} \qquad (4.4\text{-}1)$$

Incidence in yaw $\beta = \tan^{-1} v/U$ and for small angles $\beta \simeq v/U$.

If the body direction is ψ_{my} and the flight path direction ψ_{fy} then

$$\dot{\psi}_{my} = r$$

$$\text{and } \psi_{fy} = \psi_{my} + \beta$$

$$\dot{\psi}_{fy} = r + \dot{\beta}$$

$$\text{i.e. } \dot{\psi}_{fy} = r + \dot{v}/U$$

$$\text{and } U\dot{\psi}_{fy} = Ur + \dot{v} \qquad (4.4\text{-}2)$$

The reader should note the curious difference in signs in equations (4.4-1) and (4.4-2). The fact of the matter is that there is no symmetry about a

right-handed set of axes!

Consider now the motion of a particle whose velocity is U and whose velocity vector is rotating at a rate $\dot{\psi}_f$. We know that the acceleration is directed towards the centre of rotation and is given by

$$f = \dot{\psi}_f^2 . R \text{ where } R \text{ is the radius of turn.}$$

$$\text{But } U = \dot{\psi}_f R$$

$$\therefore f = U\dot{\psi}_f \qquad (4.4\text{-}3)$$

Denoting acceleration along the y axis f_y and along the z axis f_z and substituting (4.4-3) in (4.4-1) and (4.4-2) we have:

$$f_y = \dot{v} + rU$$

$$\text{and } f_z = \dot{w} - qU$$

$$\text{or } f_y m = m\,(\dot{v} + rU) = Y \qquad (4.4\text{-}4)$$

$$\text{and } f_z m = m\,(\dot{w} - qU) = Z \qquad (4.4\text{-}5)$$

which are Euler's equations with no cross-coupling c.f. (4.2-8) and (4.2-9). These equations are saying that flight path rate comprises body rate and incidence rate (v and w are proportional to incidence); and it follows that when the incidence is constant and the missile is executing a steady turn, then body rate and trajectory rate are numerically the same. In control language one often says that in the steady state body rate is the same as trajectory rate.

4.5 AERODYNAMIC DERIVATIVES

Aerodynamic derivatives are devices enabling control engineers to obtain transfer functions defining the response of a missile to aileron, elevator or rudder inputs. The method is not new; an analysis of the torque-speed characteristic of a hydraulic motor controlled by a piston valve reveals that the system is non-linear in that the output-input relationship is input amplitude dependant. Unfortunately there is no general solution to any non-linear differential equation. Yet we can make a fairly accurate estimate of performance, certainly from the stability aspect by linearising the equations and taking small perturbations about a given operating point to obtain the slopes of the curves.

Consider now the graphs shown in Fig. 4.5-1, which show rolling moment $L_{(\xi)}$ as a function of aileron angle ξ for a particular cruciform missile with rear controls for sea level at $M = 1.9$. $L_{(\xi)}$ is not a linear function of ξ for two reasons:

(a) aileron effectiveness decreases with total incidence θ

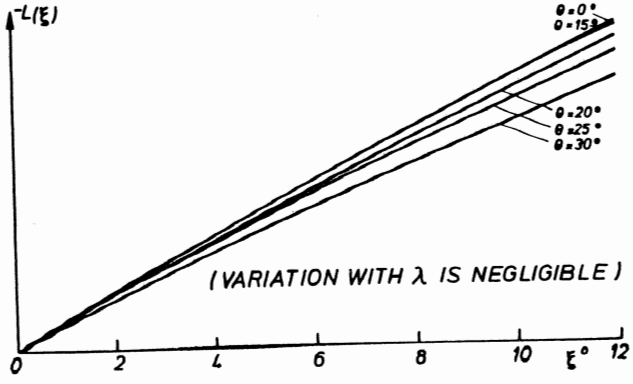


FIG. 4.5–1 Aileron effectiveness as a function of aileron angle

(b) for a given θ, L_ξ is not a linear function of ξ, although the graph does at least pass through the origin.

L_ξ is defined as:

$$L_\xi = \frac{\partial L(\xi)}{\partial \xi}, \text{ where } \xi \text{ is in radians}$$

The incremental moment L due to a small increment ξ is therefore given by

$$L = L_\xi \, \xi$$

where the value of L_ξ depends on the operating point. Bearing in mind that in most applications ξ is unlikely to exceed a few degrees we might reasonably regard L_ξ as a constant; whatever our assumptions the concept of aerodynamic derivatives enables us to model the missile response on an analogue hybrid or digital computer.

L_p is the damping derivative in roll and has dimensions of torque/unit roll rate. Since the torque will always oppose the roll motion its algebraic sign is invariably -ve. This derivative is often regarded as a constant for a given Mach number (or speed if subsonic) and operating height. There are no other important derivatives in roll.

We now consider the pitch and yaw derivatives. The normal force on a body due to pitch incidence is usually expressed as:

$$\tfrac{1}{2} \rho U^2 \, S \, C_z$$

where ρ is the air density, S is a reference area, usually the maximum body cross-section and C_z is called the normal force coefficient which is a function of incidence for a given Mach number and height. For a symmetrical missile the normal force coefficient C_y associated with incidence in yaw has

the same value as C_z. The following derivatives are defined

$$Y_\beta = \frac{\partial C_y}{\partial \beta} \cdot \tfrac{1}{2} \rho U^2 S \text{ and } Z_\alpha = \frac{\partial C_z}{\partial \alpha} \cdot \tfrac{1}{2} \rho U^2 S$$

It is important to note that these lateral force derivatives are calculated from the total force from the wings, body and control surfaces on the assumption that the control surfaces are in the central position. Most wings and control surfaces are designed to produce a normal force proportional to incidence (provided the incidence is not too large) but the normal force from a slender body includes a term proportional to incidence and another to (incidence)^2. This effect is discernible for a typical supersonic missile with rear controls in Fig. 4.5-2.

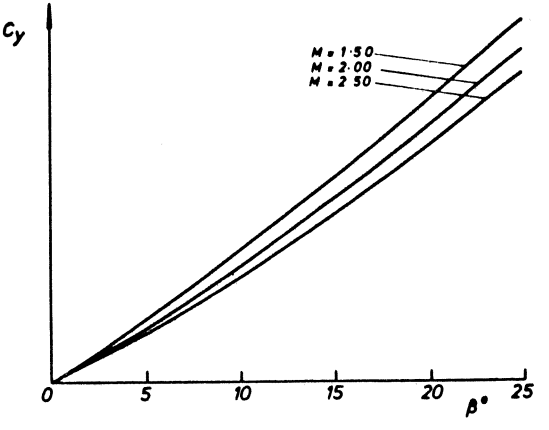


FIG. 4.5-2 Normal force coefficient as a function of incidence

There can also be variations of normal force with λ (Fig 4.1-1) but for a given Mach number and height Y_β and Z_α are generally only weakly dependant on incidence. It is often convenient to divide these two derivatives by the missile velocity U; they then become Y_v and Z_w respectively. Since for small angles $v/U \simeq \beta$ and $w/U \simeq \alpha$ the following relationships hold.

$$Y_v v = Y_\beta \beta \text{ i.e. } Y_v = Y_\beta / U \qquad (4.5\text{-}1)$$

$$\text{and } Z_w w = Z_\alpha \alpha \text{ i.e. } Z_w = Z_\alpha / U \qquad (4.5\text{-}2)$$

Y_ζ is the force derivative due to rudder deflection. For canard controls this shows little variation with incidence but for rear controls there is usually a small loss of control force with incidence due to the downwash from the wings but this is rarely more than about 15% for body incidences 15^o or so. It is important to note that if the body incidence is β and the rudder

deflection is ζ then the actual rudder incidence is $\beta + \zeta$; nevertheless the total force increment is not $Y_\beta\beta + Y_\zeta(\beta + \zeta)$ but $Y_\beta\beta + Y_\zeta\zeta$ since the normal force due to the undeflected rudder has already been included in the estimation of Y_β.

Y_r is a difficult one. It is the lateral force derivative due to unit yaw rate. It is difficult to calculate, difficult to measure in a wind tunnel and fortunately turns out to be very small and not significant.

N_v is the force derivative Y_v times the distance of the c.p. from the c.g.; this distance is known as the static margin. It is essentially an aerodynamic stiffness term and is a measure of the stability of the missile. If the c.p. is behind the c.g. any perturbation of the missile in yaw incidence will result in an aerodynamic restoring moment. Conversely if the c.p. is in front of the c.g. any perturbation of the missile in incidence will induce more incidence; the missile is said to be statically unstable. Fig 4.5-3 shows the shift in c.p. for a typical supersonic missile with rear controls.

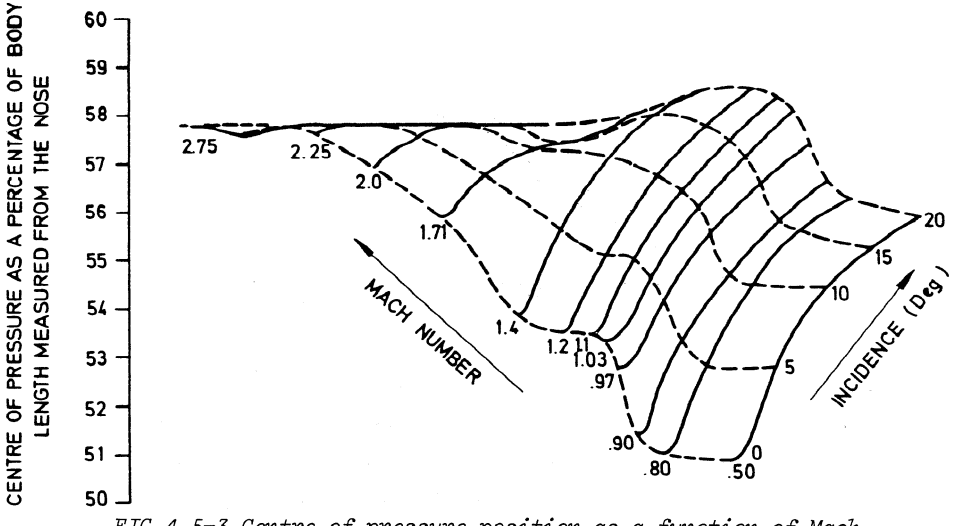


FIG 4.5-3 Centre of pressure position as a function of Mach number and incidence ($\eta = \zeta = 0^o$; $\lambda = 0^o$ *or* 90^o)

If the c.g. is about 50% of the length of the missile aft of the nose one would expect the c.p. to be typically 50-55% of the missile length aft of the nose. It is seen that at subsonic speeds and very low supersonic speeds the c.p. tends to be rather more forward than at higher Mach numbers. Also, the changes in c.p. with incidence can be considerable at low speeds; this is mainly due to the general rearwards shift in the c.p. of the body with increasing incidence, the c.p. of the control surfaces and wings changing very little. Unfortunately, the position of the c.p. is also a function of the incidence

plane angle λ. For a symmetrical missile the picture is identical to Fig 4.5-3 for $\lambda = 0^{o}$ and $\lambda = 90^{o}$ but it is slightly different for $\lambda = 45^{o}$ say.

N_{ζ} is the moment derivative due to rudder deflection and is Y_{ζ} times the distance of the c.p. of the rudder from the c.g. If Y_{ζ} is sensibly constant then N_{ζ} will vary only if the c.g. moves.

N_{r} is the damping derivative in yaw and is the aerodynamic moment per unit rate of yaw. It is a small term and is not sensitive to incidence.

We have now discussed the main roll and yaw derivatives. There is another set for the pitch plane and for a symmetrical missile these will be identical in magnitude to the yaw derivatives. There are some changes in algebraic signs as we go from the yaw plane to the pitch plane but this is due to the essential assymetry of an orthogonal right-handed set of axes; this has already been noted in equations 4.4-1 and 4.4-2.

Fig 4.5-4 provides a useful general picture of how the force derivative for the body + wings + controls varies over a wide range of speeds. The missile is conventional in shape with low aspect-ratio wings. The actual values of C_{z} will depend on the shape of the missile and the actual reference area selected. The assumption that normal force is proportional to incidence is seen to be a fair approximation especially at the higher speeds.

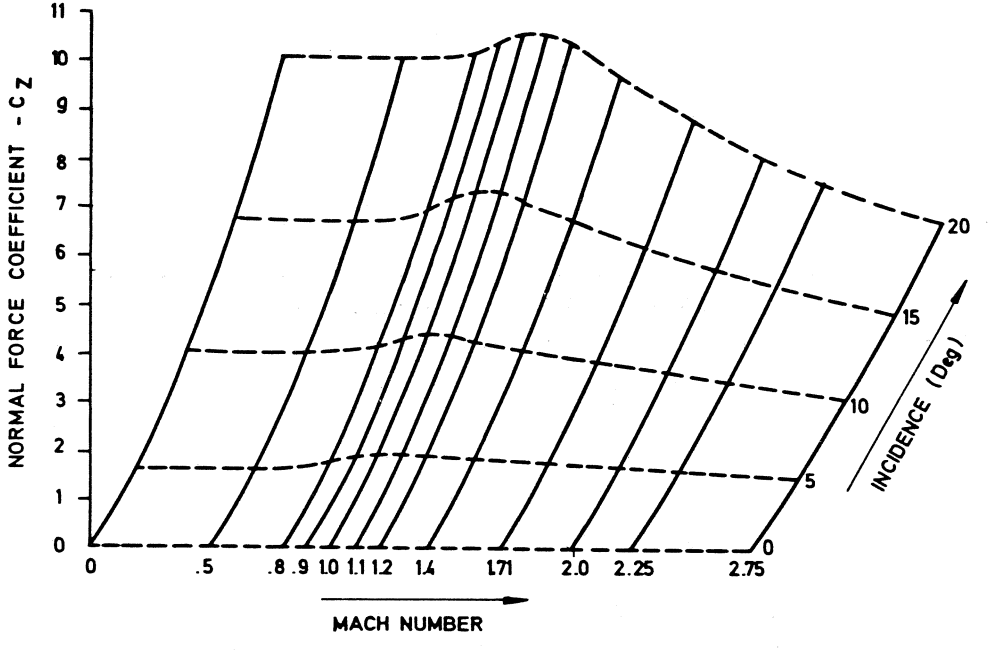


FIG 4.5-4 Normal force coefficient as a function of Mach number and incidence ($\eta = \zeta = 0^{o}$; $\lambda = 0^{o}$ *or* 90^{o})

For control purposes it is convenient to divide the force derivatives by the mass and the moment derivatives by the respective moment of inertia. It will be seen from Table 4.5-1 that with the exception of z_η, y_ζ, m_w and n_v the derivatives are then in semi non-dimensional form, (z_q and y_r are of no practical significance). M_w and N_v change sign for an unstable missile and M_η and N_ζ are of opposite sign for canard controls.

TABLE 4.5-1 ALGEBRAIC SIGNS AND DIMENSIONS OF THE MAIN AERODYNAMIC DERIVATIVES

Symbol			Algebraic sign	Dimensions	Semi non-dimensional form	
Roll	Pitch	Yaw			Symbol	Dimensions
L_ξ			$-ve$	Nm	$\ell_\xi = L_\xi/A$	sec^{-2}
L_p			$-ve$	Nm sec^{-1}	$\ell_p = L_p/A$	sec^{-3}
	Z_w		$-ve$	Nm^{-1} sec	$z_w = Z_w/m$	sec^{-1}
		Y_v	$-ve$	Nm^{-1} sec	$y_v = Y_v/m$	sec^{-1}
	Z_η		$-ve$	N	$z_\eta = Z_\eta/m$	m sec^{-2}
		Y_ζ	$+ve$	N	$y_\zeta = Y_\zeta/m$	m sec^{-2}
	Z_q		*usually* $-ve$	N sec	$z_q = Z_q/m$	m sec^{-1}
		Y_r	*usually* $+ve$	N sec	$y_r = Y_r/m$	m sec^{-1}
	M_w		$-ve$	N sec	$m_w = M_w/B$	m^{-1} sec^{-1}
		N_v	$+ve$	N sec	$n_v = N_v/C$	m^{-1} sec^{-1}
	M_η		$-ve$	Nm	$m_\eta = M_\eta/B$	sec^{-2}
		N_ζ	$-ve$	Nm	$n_\zeta = N_\zeta/C$	sec^{-2}
	M_q		$-ve$	Nm sec	$m_q = M_q/B$	sec^{-1}
		N_r	$-ve$	Nm sec	$n_r = N_r/C$	sec^{-1}

4.6 AERODYNAMIC TRANSFER FUNCTIONS

At the commencement of a desk study design of a missile control system one must make certain assumptions. A traditional assumption is one of the linearity of the hardware i.e. the electronics, fin servos, instruments (if any) and the equations of motion. Indeed, as has already been remarked, linearity is a necessary constraint for one to use operator methods to analyse the system response. However, linearity of the equations of motion imply linearity of the aerodynamics - a defendable tenet when the concept of small perturbations is adhered to but on less certain grounds when the missile is exercised through incidence changes of 20^o or more and the attendant body rates are large.

The general method of designing a missile control system is this. One considers a typical speed (or Mach number if the missile is supersonic) and height the missile will operate at. One can take a set of aerodynamic derivatives for zero incidence and design a control system on the assumption that the missile is exercised through small perturbations about zero incidence; in other words one regards the aerodynamic derivatives as constants within this small range. Since the control system is designed to meet a specification concerning steady state gain, bandwidth, phase lag and damping, we investigate the response as a function of frequency on these assumptions. We now look at the derivatives at 5° incidence (say) and the same calculations are carried out to see if the specification can be met with these new values of derivatives. The calculations are repeated for as many combinations of incidence and roll angle as judgment deems advisable. And finally the missile may well have to operate over a range of speeds and heights. A representative number of these operating conditions is also tested. When one considers that the inertia, mass and position of the c.g. may vary due to propellant usage it is seen that the amount of detailed work involved in checking and possibly modifying the design can be considerable. The conclusion one draws on completion of this exercise is that if the design is satisfactory at all these points then it is satisfactory at all intermediate points.

The linearised approach can now be taken further. The guidance and control systems can be put together on a relatively simple computer and a preliminary assessment of system performance can be made in one plane. The final simulation on a digital or hybrid computer will almost certainly programme one or two of the important derivatives as a function of incidence and will take note of any cross-coupling that may occur.

We will now write down the control equations for a missile, omitting the force equation along the x axis as this neither affects the roll, pitch or yaw motion. We will also omit the gravity term and consider forces and moments which are purely of aerodynamic origin. The effect of gravity will be considered as a separate issue under autopilot design. Equations (4.2-9) to (4.2-13) can be written

$$mf_y = m(\dot{v} + rU) = Y = Y_v v + Y_r r + Y_\zeta \zeta$$

$$\text{i.e.} \quad f_y = \dot{v} + rU = y_v v + y_r r + y_\zeta \zeta \tag{4.6-1}$$

$$mf_z = m(\dot{w} - qU) = Z = Z_w w + Z_q q + Z_\eta \eta$$

$$\text{i.e.} \quad f_z = \dot{w} - qU = z_w w + z_q q + z_\eta \eta \tag{4.6-2}$$

$$A\dot{p} = L = L_p p + L_\xi \xi \quad (\textit{assuming } B \simeq C)$$

$$\text{i.e. } \dot{p} = \ell_p p + \ell_\xi \xi \tag{4.6-3}$$

$$B\dot{q} = M = M_w w + M_q q + M_\eta \eta$$

$$\text{i.e. } \dot{q} = m_w w + m_q q + m_\eta \eta \tag{4.6-4}$$

$$C\dot{r} = N = N_v v + N_r r + N_\zeta \zeta$$

$$\text{i.e. } \dot{r} = n_v v + n_r r + n_\zeta \zeta \tag{4.6-5}$$

Roll rate/aileron p/ξ

This is the simplest aerodynamic transfer function. Equation 4.6-3 can be rewritten:

$$\dot{p} - \ell_p p = \ell_\xi \xi$$

or in the transfer function form $\dfrac{p}{\xi} = \dfrac{\ell_\xi}{s - \ell_p} = \dfrac{-\ell_\xi/\ell_p}{T_a s + 1}$ (4.6-6)

where $-\ell_\xi/\ell_p$ can be regarded as a steady state gain and $T_a = 1/-\ell_p$ and can be regarded as an aerodynamic time constant.

Lateral acceleration/rudder f_y/ζ

Equations (4.4-2), (4.6-1) and (4.6-5) can be used to eliminate r and v to yield:

$$\frac{f_y}{\zeta} = \frac{y_\zeta s^2 - y_\zeta n_r s - U(n_\zeta y_v - n_v y_\zeta)}{s^2 - (y_v + n_r)s + y_v n_r + U n_v} \tag{4.6-7}$$

The reader will notice that y_r does not appear in this transfer function; it has already been remarked that it is a very small quantity and in this context it is usually omitted. Some consideration of the individual terms is worthwhile. Clearly $(y_v n_r + U n_v)$ is (the undamped natural frequency)^2 and is usually referred to as the weathercock frequency; it reflects the tendency for a stable missile to return to the unperturbed zero incidence position. If, for a moment, we regard the missile as a weathercock pinned at its c.g. then the undamped natural frequency is given by:

$$\omega_n^2 = \frac{\text{the restoring moment/unit angular deflection}}{\text{moment of inertia about c.g.}}$$

$$= \frac{N_\beta}{C} = \frac{UN_v}{C} = U n_v$$

This does not agree with the expression given in (4.6-7) since there is an extra term $y_v n_r$. This is not surprising since a missile is not pinned at its c.g. and we would expect to see additional moments opposing the motion of a body with six degrees of freedom. Nevertheless in most cases when the static margin is from 2-5% of the missile length we find that $U n_v$ is at least twenty

times the magnitude of $y_v n_r$.

Consider now a typical surface-to-air missile with rear controls whose main yaw derivatives for M = 1.4 and height 1500 m (U = 467 m/sec) are:

y_v	n_v	y_ζ	n_ζ	n_r
-2.74	0.309	197	-534	-2.89

The missile is 2 m long and C = 13.8 kgm^2 and m = 53 kg. Hence the static margin = $N_v/Y_v = n_v C/y_v m$ = 0.029m or about 1.5% of the length. Inserting the values for the derivatives in (4.6-7) we obtain:

$$\frac{f_y}{\zeta} = \frac{197s^2 + 570s - 467\ (1460 - 60.8)}{s^2 + (2.74 + 2.89)s + (7.9 + 144)}$$

Thus ω_n^2 = 7.9 + 144 and ω_n = 12.4 rad/sec; we note $\omega_n^2 \simeq Un_v$. The damping ratio μ is given by $2\mu\omega_n$ = 2.74 + 2.89 and hence μ = 0.23. This damping ratio incidentally will decrease as the square root of the air density and therefore can be very low at high altitudes; the damping terms vary with air density, and so do the force and moment terms, but the inertia remains the same. The steady state gain is - 467 x 1399/152 = - 4300. This means that if the aerodynamics were linear 0.1 rad rudder deflection would produce a lateral acceleration in the $-y$ direction of 430 m/sec^2 or about 43g. This appears to be a very high aerodynamic gain, but there is often a good reason for this as will be discussed under autopilot design. Now a second order response (surely the airframe is yet another example of a damped spring-mass system) is completely defined by the steady state gain, the undamped natural frequency and the damping ratio. What then is the significance of the extra two numerator terms? In the time domain all they can do is to define the initial values at t = 0+ (as opposed to the initial conditions at t = 0-). If we consider a unit step input then the lateral acceleration at t = 0+ is given by the force from the rudder only divided by the missile mass i.e. $f_y = y_\zeta$. Hence the value at t = 0+ is given by the coefficient of s^2 in the numerator; the coefficient of s^2 in the denominator is unity. This result and that for the initial value of the slope can be obtained from the initial value theorem. The numerator can be written:

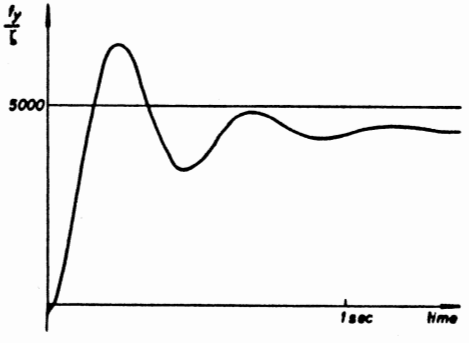
$$-655{,}000\ (-\ .000301s^2 - .00087s + 1)$$

$$\text{this is: } -655{,}000\ (s/59.1 + 1)(-s/56.2 + 1)$$

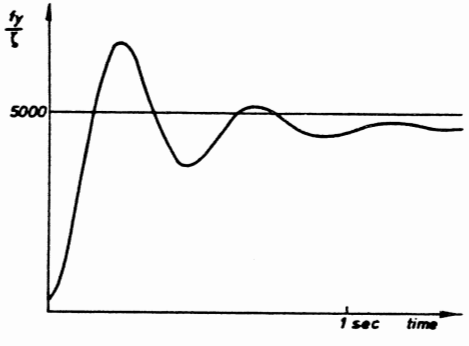
This is saying that in the frequency domain there are two zeros, one non-minimum phase, of about the same value, whose phase contribution will approximately cancel over a range of frequencies but whose amplitude contribution will become reckonable in the range 60 rad/sec and above. These terms

can be a nuisance in autopilot design as they prevent the continuous attenuation with frequency associated with a simple second order system. The reason is obvious; all the time the rudder is moving there will be a lateral acceleration whatever the frequency.

What happens if we take the controls to the front of the missile and move the wings back so that the position of the c.p. remains unchanged? Assume all aerodynamic derivatives are unchanged numerically; there will be a change in the algebraic sign of n_ζ i.e. it is now +ve. This will have the effect of changing one term in the numerator from -467 (1460 - 60.8) to -467 (-1460 - 60.8) = 710,000. The steady state gain is now 710, 000/152 = 4670 an increase of about 9%. The reason for the change in algebraic sign of the steady state gain is due to the fact that +ve canard rudder deflection produces +ve incidence. The reason for the increase in gain is due to the fact that canards, wings and body are all producing normal force in the same sense. Figures 4.6-1 (a) and (b) show the time responses for f_y, for a step rudder input for the tail controlled missile compared with a similar missile with canards. For convenience of comparison both responses are regarded as +ve.



(a) Rear controls



(b) Canard controls

FIG. 4.6-1 Lateral acceleration for a step rudder input

Finally, consider the original rear controlled missile whose c.g. has moved forward slightly to increase the small static margin by a factor of four. The steady state gain is decreased approximately by a factor of four, the weathercock frequency is doubled and the damping ratio is halved. The significance of the static margin should now be apparent. A large static margin results in:

(a) a small steady state gain
(b) a high weathercock frequency
(c) a very low damping ratio.

A small static margin results in:

(a) a large steady state gain
(b) a low weathercock frequency
(c) a low or moderate damping ratio.

It follows therefore that the position of the c.p. and the c.g. are of vital interest to the missile control system designer.

For a symmetrical missile the transfer function f_z/η must be essentially the same as for f_y/ζ. There are some changes in sign however. For instance +ve rear elevator deflection eventually produces acceleration in the +ve z direction.

Rate of change of flight path/rudder $\dot{\psi}_{fy}/\zeta$

Flight path rate and lateral acceleration are essentially the same things. There is a difference in scale only. To obtain flight path rate we divide f_y by U.

Body rate/rudder r/ζ

To obtain this transfer function $\dot{\psi}_{fy}$ and v must be eliminated from equations (4.4-2), (4.6-1), and (4.6-5) to yield:

$$\frac{r}{\zeta} = \frac{n_\zeta s - n_\zeta y_v + n_v y_\zeta}{s^2 - (y_v + n_r)s + y_v n_r + U n_v} \qquad (4.6\text{-}8)$$

If one compares this with (4.6-7) one sees that the undamped natural frequency and the damping ratio are the same; indeed, flight path rate, lateral acceleration, body rate and incidence are all oscillatory modes of the missile and these modes are identical in frequency and damping. Only the initial values, slopes and steady state gains are different. Using the same set of aerodynamic derivatives the steady state gain is therefore -1399/152 = -92 which is the same as that for $\dot{\psi}_{fy}/\zeta$. Body rate has the same steady state value as flight path rate. What then is the essential difference between body rate and flight path rate? Consider the numerator in equation (4.6-8) which can

be written

$$-(n_\zeta y_v - n_v y_\zeta)(T_i s + 1)$$

where T_i is usually called the incidence lag (most confusing!) and has the value:

$$T_i = \frac{n_\zeta}{-n_\zeta y_v + n_v y_\zeta} = \frac{1}{-y_v\,(1 - n_v y_\zeta / n_\zeta y_v)} \qquad (4.6\text{-}9)$$

$$= \frac{1}{-y_v\,(1 - x^*/\ell_c)} \simeq \frac{1}{-y_v} = .365 \text{ secs}$$

where ℓ_c is the distance from the c.g. to the c.p. of the rudder and x^* is the static margin.

It is seen that, for a given missile this time constant increases with altitude and increases if the speed falls. It tends to be a large time constant and typically varies from 0.25 to 2 sec. Neglecting the small numerator terms in the transfer function for f_y/ζ, the relationship between lateral acceleration, flight path rate and body rate is shown in Fig. 4.6-2.

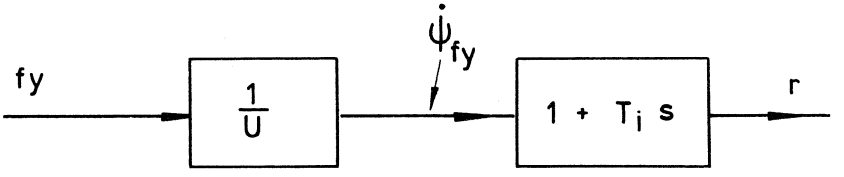


FIG. 4.6-2 The approximate relationship between lateral acceleration, flight path rate and body rate

It is useful to regard body rate as phase advanced flight path rate; or to a different scale body rate is phase advanced lateral acceleration. This is the reason why feedback from a rate gyro in an autopilot whether it is employed in an aircraft, helicopter or guided missile tends to damp the weathercock mode; it contains a powerful derivative term. Figure 4.6-3 shows body rate for a step rudder input for the previous missile with rear controls.

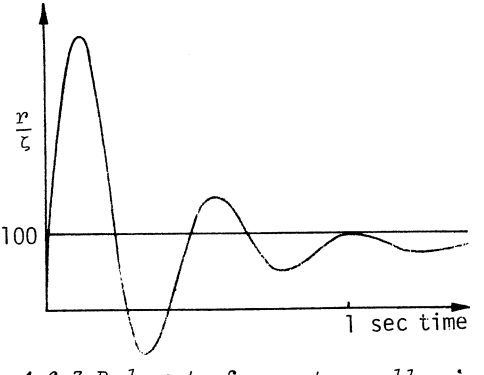


FIG. 4.6-3 Body rate for a step rudder input

The concept of body rate and flight path rate can be confusing. Consider Fig. 4.6-4(a). A missile is flying straight and the rudder is suddenly deflected as shown. In Fig. 4.6-4(b) the missile is shown having turned with a steady state incidence equal to $\psi_{my} - \psi_{fy}$. If the incidence is constant

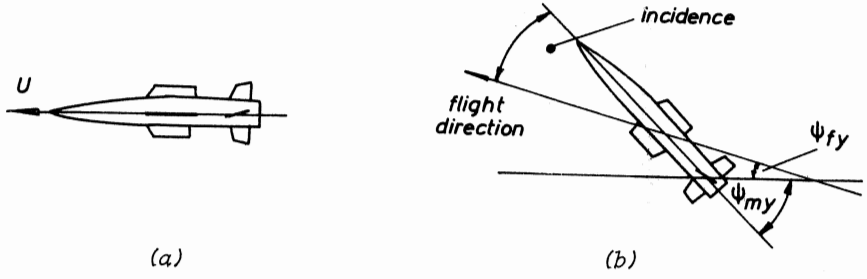


FIG. 4.6-4 Flight path direction lags behind body direction

then the body rate must be equal to the flight path rate; and yet in the same time the angle the body has turned through has been greater than the change in flight path angle. The explanation of this is that during the transient period when the body incidence was building up from zero the body rate must have been faster than the flight path rate. How else could it have got ahead? In other words during the transient period body rate was equal to flight path plus its differential times a constant.

Body incidence/rudder β/ζ

To obtain this transfer function the substitution $\beta = v/U$ is made and $\dot{\psi}_{fy}$ and r are eliminated in the same three equations as before to yield:

$$\frac{\beta}{\zeta} = \frac{(y_\zeta s - n_\zeta U - y_\zeta n_r)/U}{s^2 - (y_v + n_r)s + y_v n_r + U n_v} \qquad (4.6\text{-}10)$$

The denominator has not changed of course. The numerator term in "s" is insignificant and the steady state gain is 3.52 rad/rad i.e. 1^o rudder will produce 3.52^o body incidence in the steady state. This is obtained by having a powerful rudder with a long moment arm and a small static margin. Figure 4.6-5 shows the incidence response of the missile to a unit step rudder input. This is indistinguishable from a simple second order response.

4.7 ALTITUDE AND SPEED CONVERSION FACTORS FOR AERODYNAMIC DERIVATIVES

Subsonic conditions

Supposing aerodynamic derivatives are given for standard conditions at sea level and for a given speed say 100m/sec. How do we convert a given derivative

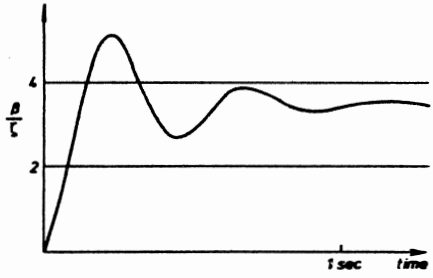


FIG. 4.6–5 Body incidence for a step rudder input

to say 150 m/sec and a given height? Neglecting effects due to Reynolds Number which are usually small or very small we can say that all normal forces are proportional to $\rho U^2/2$. Therefore the force derivative Y_ζ and the moment derivatives N_ζ and N_r are very simply converted by the ratio $\rho U^2/\rho_0 U_0^2$. However N_v and Y_v must be approached with caution remembering that Y_v is the lateral force developed per unit side-slip velocity and that $\beta = v/U$.

$$\text{Actual force} = Y_\beta \beta = Y_v v \text{ i.e. } Y_v = Y_\beta/U$$

if we know Y_v for one condition we multiply it by $\rho U/\rho_0 U_0$ (not the square of the velocities). The same goes for N_v.

Supersonic conditions

In supersonic conditions Mach number is now an independent variable i.e. normal force is now a function of M and proportional to $\rho U^2/2$. For a given M:

$$\text{Normal force} \propto \rho U^2/2 \propto \rho M^2 a^2/2 \text{ where } a \text{ is the local speed of sound}$$

$$\propto \frac{\rho}{\rho_0} \left(\frac{a}{a_0}\right)^2 \times \frac{1}{2} \rho_0 a_0^2 M^2$$

Table 4.7-1 shows conversion factors for the main derivatives on the assumption that we know them for a given Mach number at sea level.

TABLE 4.7-1 ALTITUDE CONVERSION FACTORS FOR AERODYNAMIC DERIVATIVES FOR A GIVEN MACH NUMBER

Derivatives	Altitude Conversion Factor	Values of factors at altitude (m)				
		0	5000	10000	15,000	20,000
*Y_v N_v	$\frac{a}{\rho_0 a_0}$	1.0	0.567	0.296	0.1355	0.0615
*Y_ζ N_ζ Y_r N_r L_p L_ξ	$\frac{\rho}{\rho_0} \left(\frac{a}{a_0}\right)^2$	1.0	0.536	0.262	0.1168	0.0531

*Likewise for the corresponding pitch derivatives.

CHAPTER 5

MISSILE INSTRUMENTS

5.1 INTRODUCTION

Most closed loop systems incorporate an instrument (or transducer as they are often called) to measure the controlled quantity. The output from the transducer is usually a voltage and this is readily compared with the input or reference if this also exists as a voltage. Obvious examples are potentiometers and tachogenerators to monitor the position and speed of a shaft. Consider now any aspect of the motion of a missile through space. Forces and moments will produce accelerations and hence velocities and displacements with respect to the earth; or, as is often stated with respect to inertial space. If we wish to make a closed loop system of the means of controlling the motion of a missile then instruments will be required to measure accelerations, velocities and displacements in space. Clearly, conventional potentiometers and tachogenerators do not do this. Instead, accelerometers, rate gyros and position gyros are used. It is usual to call a system comprising missile fin or thrust vector servos, an airframe, instruments and any electronics and networks necessary to close the loop an autopilot; but this nomenclature is not universal. In this chapter, relatively cheap instruments which are used in short and medium range missiles are discussed. The very sophisticated instruments used in inertial navigation and inertially guided missiles are not discussed since this forms part of a completely different subject which is outside the scope of this book.

5.2 ELEMENTARY THEORY OF GYROSCOPES

We all know that if a heavy rotor is suspended in a set of friction-free bearings and is spun up it tends to maintain its spin axis in the same direction in space. We realise that if the gyro is moved to another point on the earth's surface it will appear to point in another direction due to the curvature of the earth's surface. Also, its spin axis will appear to change in direction due to the rotation of the earth. Since we are interested in times of flight typically of 10-20 seconds and very occasionally up to about 100 seconds, the distances and times involved result in these effects being negligible.

Unfortunately the spin direction will wander due to unwanted moments acting on the suspended mass. Consider now a body of polar moment of inertia J spinning at a rate ω in the direction shown in Fig 5.2-1.

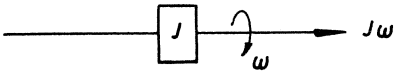


FIG. 5.2-1 Angular momentum shown as a vector

Angular velocity and angular momentum are both represented by a line perpendicular to the plane of the spin and a right handed corkscrew convention is used. If now a moment is applied in the plane of the spin the angular momentum changes magnitude but its direction is unaltered. Suppose a moment T is applied for a short time δt at right angles to the spin vector as shown in Fig 5.2-2.

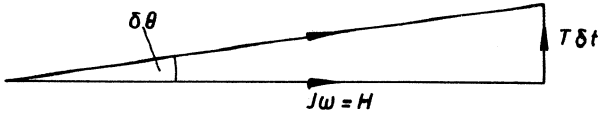


FIG. 5.2-2 Precession of rotating body

If the body rotates through $\delta\theta$ then $T\delta t$ must equal the change in angular momentum. If we denote angular momentum $J\omega = H$, then

$$H\delta\theta \simeq T\delta t$$

and in the limit

$$\Omega = \frac{d\theta}{dt} = \frac{T}{H} \qquad (5.2\text{-}1)$$

This change in direction of the spin vector is usually known as precession and it is now clear that if it is to be small moments must be kept small; a large angular momentum is an advantage.

Now consider three mutually perpendicular axes ox, oy and oz. If H is along ox (i.e. in the plane yz), and the moment is applied about oy (i.e. in the plane xz), then precession will occur about oz (i.e. in the plane xy). Equally we can say that if a body is spinning about ox and is forcibly made to precess about oz then a torque reaction will occur about oy. And finally, it is now obvious that in the absence of any moments about oy and oz the body will not precess but will remain spinning in the same direction. Before looking at constructional details two other mechanical phenomena are worthy of note.

Consider now a body of mass m suspended between two points such that the stiffness of suspension through the centroid and perpendicular to one principal

axis yy is s_y and about the other principal axis zz is s_z. If a force P acts on the body through the centroid perpendicular to yy then the deflection is P/s_y and is in the same direction as P. Similarly if a force Q acts perpendicular to zz then the deflection is in the same direction as Q and is equal to Q/s_z. See Fig 5.2-3. If now these two forces act together then

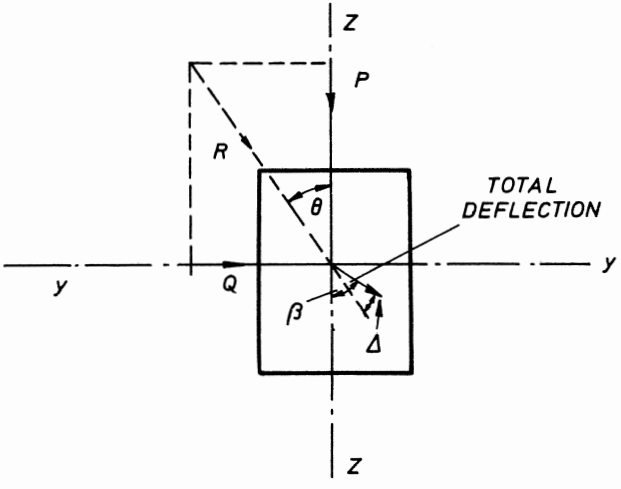


FIG. 5.2-3 Deflection due to anisoelasticity

the angle θ that the combined force R makes with the zz axis is given by

$$\tan \theta = Q/P \tag{5.2-2}$$

The angle β that the deflection makes with this axis is however:

$$\tan \beta = \frac{Q}{s_z} \Big/ \frac{P}{s_y} \tag{5.2-3}$$

and these two angles are equal only if $s_z = s_y$. The general condition therefore is for deflections not to be in the same direction as the force. The deflection has two components, one in the same direction as the applied force and one perpendicular to it. Call this latter component Δ. The moment about the centroid is therefore $R\Delta$ and since $\Delta \propto R$, the moment $\propto R^2$. If the sense of R is reversed the sense of Δ is also reversed and therefore the sense of the moment is unaltered. An example of a reversing force is the inertia force due to vibration. Gyro designers say that "anisoelasticity rectifies the effect of vibration". A partial answer to this problem is to make the suspension as stiff as possible but the perfect design has the two principal bending axes equally stiff; it is then found that it is equally stiff in all directions. In practice most gyros used in tactical missiles

will have a wander rate proportional to g^2 due to anisoelasticity. Finally, the phenomena of nutation must be mentioned. Nutation is a free oscillation of a gyro rotor in its gimbals and the frequency can be determined by setting the external moments to zero. The inner gimbal is at an angle θ to the orthogonal position. Assume the angular momentum H of the rotor is constant. Let the inertia of the rotor plus inner gimbal about oy be J_y and the rotor plus both gimbals about oz be J_z, see Fig 5.2-4.

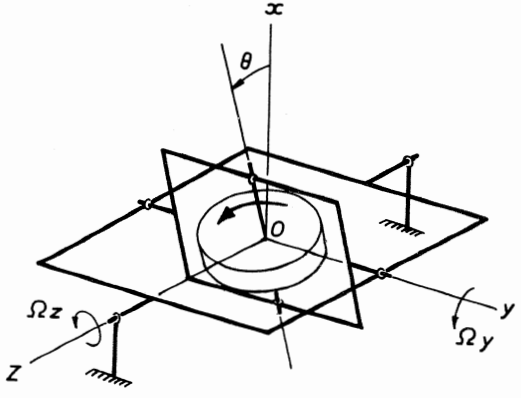


FIG. 5.2-4 Rotor rotating in gimbals

$$T_{oy} = 0 = J_y D\Omega_y - H \cos\theta\, \Omega_z \qquad (5.2\text{-}4)$$

$$\text{and } T_{oz} = 0 = J_z D\Omega_z + H \cos\theta\, \Omega_y \qquad (5.2\text{-}5)$$

It will be noted that any small damping torques have been ignored. Equation 5.2-5 can be differentiated to read:

$$-D\Omega_y = J_z D^2\Omega_z / H \cos\theta$$

and this value for $D\Omega_y$ can be substituted in equation 5.2-4 to yield:

$$\left(D^2 \frac{J_y J_z}{H^2 \cos^2\theta} + 1\right)\Omega_z = 0 \qquad (5.2\text{-}6)$$

and hence the nutation frequency $\omega_n = \frac{H \cos\theta}{\sqrt{J_y J_z}} = \omega \frac{J \cos\theta}{\sqrt{J_y J_z}}$ (5.2-7)

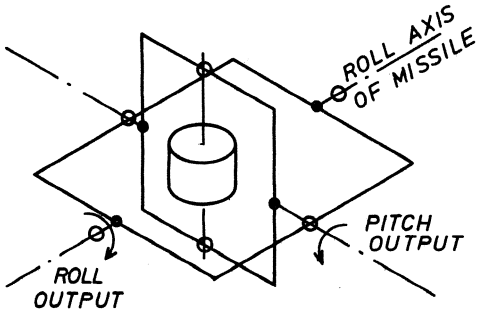
where J is the rotational inertia as before. If the rotor is a sphere then J_y must be more than J and J_z must be more than J_y. In practice one will find that the nutation frequency for $\theta = 0$ is about 70% of the rotational frequency. If now the rotor speed is 24000 rev/min then the nutation frequency is likely to be about 0.7 x 400 = 280 Hz. Fundamental bending modes of missiles 6m long are likely to be around 25 Hz and 100 Hz for small missiles about 1m long. It is essential that this mode of oscillation is not

excited and one can see that it is desirable to keep the rotor speed as high as possible and to avoid large values of θ.

5.3 FREE OR POSITION GYROS

The simplest position gyro has three degrees of freedom and is often called a "free" gyro. It consists of a heavy rotor spinning at a high speed in a light framework called a gimbal. This gimbal is supported in bearings in another gimbal ideally at right angles to the spin axis. Figs 5.3-1 to 5.3-3 show three alternative orientations of a position gyro. If one angular position transducer detects relative movement between the missile frame and outer gimbal, and another the relative movement between the two gimbals, then it is possible with one free gyro to measure two angular rotations of the missile. If three angular rotations have to be measured then two such gyros are required. It is important to note that the vehicle can have unlimited freedom to rotate about the outer gimbal axis and will always have unlimited angular freedom about the gyro spin axis without impairing the original orthogonal nature of the three axes. Consider now the yaw and pitch gyro in the configuration shown in Fig 5.3-2. Rotation of the vehicle about the pitch axis will gradually erode one of the degrees of freedom. A movement of 90^{o} will destroy one of the degrees of freedom completely and "toppling" of the gyro will result. This is because any roll of the missile will constrain the gyro to precess which will in turn cause an angular acceleration of the inner gimbal about the oy axis. If a missile carries two free gyros then it cannot have unlimited angular freedom in more than one direction.

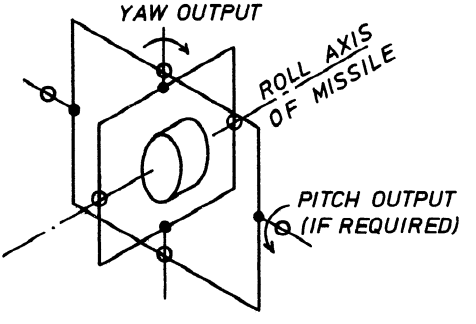
The question is now asked, if the vehicle frame rotates such that the original vehicle x, y and z axes cease to coincide with the gimbal axes, do the angles as indicated by the pick-offs still indicate the correct vehicle rotations? Reference is again made to Fig 5.3-2. The missile can clearly roll about the x axis without in any way altering the essential geometry of the system; it is merely rotating about the spinning gyro. Also rotation of the missile about the y axis or the z axis does not affect the accuracy of measurement provided these two rotations do not occur together. Assuming now that the rotor does not wander let the missile yaw through an angle ψ and then rotate through θ about its own pitch axis. What is the rotation of the outer gimbal relative to the missile framework as this will be the indicated yaw angle ψ' and what is the relative movement between inner and outer gimbals θ' as this will be the measured pitch? Resolving these angular movements



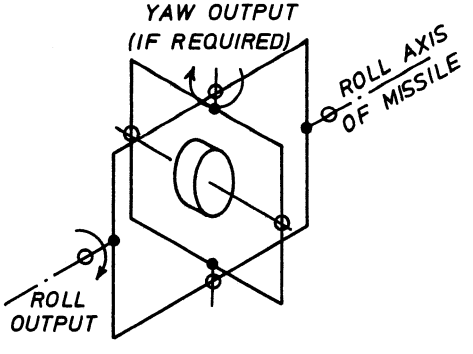
ROLL AND PITCH GYRO
(VERTICAL GYRO)

FIG 5.3-1

FIG 5.3-2



YAW (AND PITCH) GYRO
(DIRECTIONAL GYRO)



ALTERNATIVE ROLL (AND
YAW) GYRO

FIG 5.3-3

gives this result

$$\sin\theta' = \sin\theta\cos\phi \qquad (5.3\text{-}1)$$
$$\text{and } \tan\psi' = \sec\theta\tan\psi$$

Despite the apparent symmetry of the axes, these two equations are not similar. Suppose $\theta = \psi = 30^{\circ}$. These two equations tell us that the indicated pitch angle is $25\frac{1}{2}^{\circ}$ and the indicated yaw angle is $33\frac{1}{2}^{\circ}$. The same argument applies to the measurement of missile roll. One can see from Figs 5.3-1 and 5.3-3 that there are two possible orientations of the spin axis if one is interested in roll only, and depending on the anticipated subsequent missile motions one may be preferable to the other since the distortion of measurement is less. This gimbal distortion is one cause of phasing error in command systems as we shall see later.

Precision miniature ball bearings are normally used for the suspension of all three axes. Free gyros designed to run for long periods usually have air bearings to support the rotor but these are not a feature of gyros used for missiles with short times of flight. The rotors can be blast started and then allowed to coast. This method is obviously suitable for use in systems which are required to have a short total reaction time such as anti-tank, air-to-air, and short range surface-to-air systems. If compressed air is used as the power source for the missile servos then two symmetrically placed jets can be made to impinge on Pelton wheel like buckets machined around the rotor. A rotor can be spun up to 60,000 revs/min in half a second in this way. If, as is usual, the reference direction is required to be the direction at launch then the two gimbals must be uncaged immediately after the rotor is spun up. Some sort of delay mechanism is necessary. If the missile servos do not use compressed air then it is usual to blast start the rotor by means of a small separate charge of cordite. Another time delay is necessary to allow the gyros to run up before the main rocket motor is ignited. Missiles with times of flight of more than about 40 seconds have electrically driven gyros. Hysteresis motors are suitable since they have the characteristic of developing full torque from standstill up to synchronous speed. If a 400 Hz supply is available then the maximum rotor speed is 24,000 revs/min. A typical run-up time is about 20 seconds but this can be reduced to about 6 seconds if an initial over-driving technique is used. Motors are designed inside-out with an inner stator and an outside rotor. Fig 5.3-4 shows a typical blast started position gyro fitted with one potentiometer and weighing 255 grams.

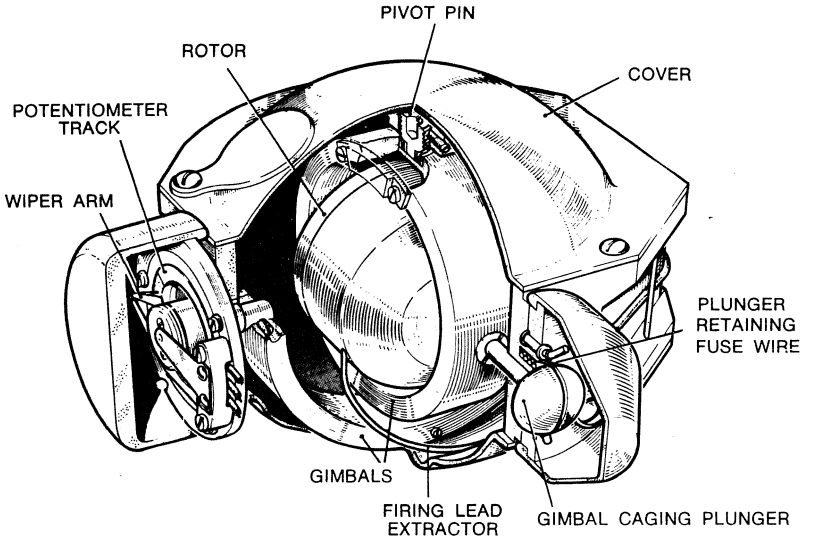


FIG. 5.3-4 Blast started position gyro

Consider now some causes of wander rates. The following particulars apply to a blast started position gyro.

Initial wheel speed:	20,000 rev/min
Wheel speed after 33 seconds:	> 75% of original speed
Weight:	0.35 kg

It is difficult to be specific about drift rates unless the environment can be defined more or less exactly. They could easily be less than 0.05^{o}/second but could approach about 0.25^{o}/second if the missile is rotating at a mean speed of 3 revs/second as the friction torques from one set of bearings will be unidirectional. Also drift rates could greatly exceed 0.05^{o}/second during high g such as the boost period. Other causes of drift are torques from slip rings, potentiometer wipers and from motor leads if the rotor is electrically driven. A two axis blast started gyro of this quality will cost about £150 at 1975 prices. Conventional gyros with lower drift rates are usually bigger and will certainly cost more.

5.4 RATE OR CONSTRAINED GYROS

The great majority of rate gyros measure angular rate about one axis. A rotor is mounted in a gimbal whose motion about an axis at right angles to the spin axis is constrained by a torsion bar or friction-free spring system. There are no other gimbals so the rotor has one degree of freedom

only about its own spin axis. Figs 5.4-1 and 5.4-2 show the mode of operation and some constructional details of a high performance single axis rate gyro. The cylindrical gimbal is enclosed in a hermetically sealed outer case and the gap between them is filled with viscous fluid in which the gimbal is floated with neutral buoyancy. The fluid provides viscous shear damping, minimal pivot friction and protection from shock. Define:

k = torsion bar stiffness

β = angular twist in the torsion bar

H = angular momentum of the rotor

J_g = total inertia about the torsion bar axis

f = viscous damping coefficient

Ω = rate of turn of missile about the input axis

If now the missile rotates about the spin axis no reaction of any sort is set up. If a steady rotation about the torsion bar axis occurs the rotor is constrained to precess at this rate also and the gyroscopic torque will have to come from direct shear of the pivot at one end and direct shear of the torsion bar at the other. If however the missile rotates about the "input axis" as indicated then precession of the rotor will occur at the expense of torsion set up in the torsion bar, and in the steady state the angle of twist will be proportional to the input rate. Equating torques about the torsion bar axis we have:

$$J_g D^2\beta + fD\beta + k\beta = \text{gyroscopic torque} = H\Omega$$

Hence

$$\frac{\beta}{\Omega} = \frac{H/k}{s^2/\omega_n^2 + 2\mu s/\omega_n + 1} \qquad (5.4\text{-}1)$$

$$\text{where } k/J_g = \omega_n^2$$

This is an example of a limited gain-bandwidth product. For a given H high gain is obtained by having a compliant torsion bar, but this only lowers the frequency response of the instrument i.e. one can have a high gain or high bandwidth but not both. For most guided missile applications a high natural frequency is desirable, usually of the order of 80 Hz or more, so a stiff torsion bar and low d.c. gain must be accepted. If, for instance ω_n = 500 rad/sec then $k/J_g = 500^2$. If the rotor speed is 24,000 rev/min i.e. 800π rad/sec then assuming the rotor inertia is approximately equal to the gimbal inertia then the d.c. gain from equation 5.4-1 is given by:

$$\frac{\beta}{\Omega} = \frac{H}{k} = \frac{J\omega}{k} = \frac{\omega}{500^2} = \frac{800\pi}{500^2} = .01 \text{ rad/rad/sec} \qquad (5.4\text{-}2)$$

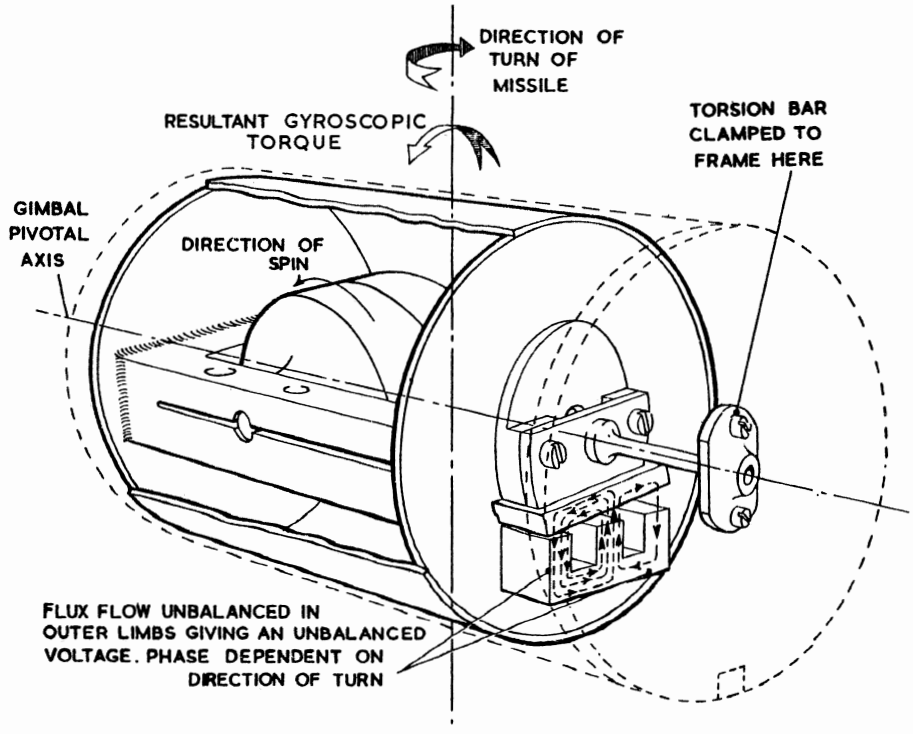


FIG. 5.4-1 Mode of operation of rate gyroscope

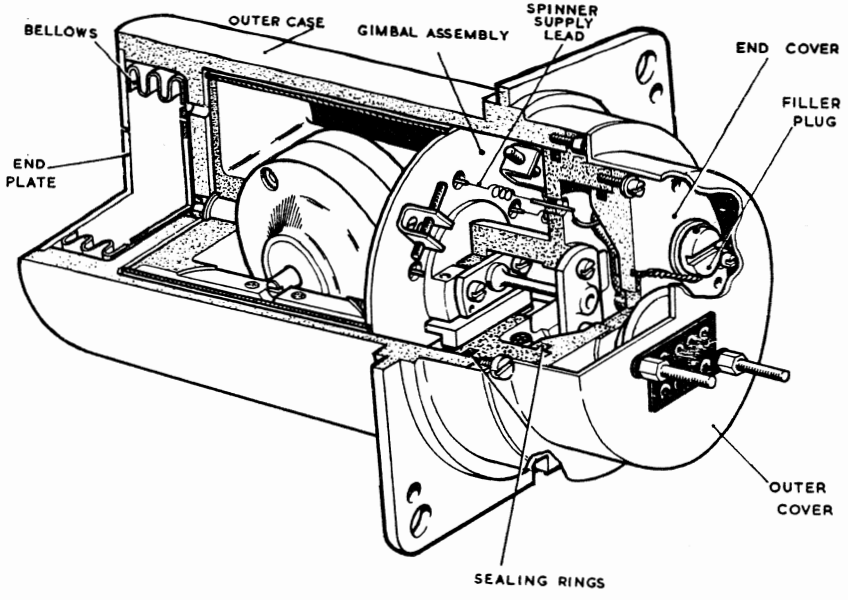


FIG. 5.4-2 Some constructional details of a rate gyroscope

If the maximum input is 4 radians/second then the maximum output is 0.04 radians or just over 2^{o}. This is why the majority of rate gyros have a.c. pick-offs for measuring these small angular movements. The one illustrated is a linear variable differential transformer often called an "E type" pick-off. If one requires a really accurate rate gyro a small angular movement is essential since any displacement of the spin axis will result in the sensitive axis being displaced by the same amount and hence cross coupling from rates in a plane at right angles to the nominal one will occur.

Very good resolution and linearity can be obtained with rate gyros but to reduce bias over a wide range of operating temperatures to much less than ± 2% of full scale calls for much skill on the part of the instrument designer (1). Rate gyros, like position gyros, can be blast started or electrically driven. If blast started and allowed to coast a continuous loss of gain with time must be accepted but if the instrument is required for autopilot damping some 20% loss of gain would not be critical. One gas fired rate gyro weighing only 130 grams costs about £100. Conversely a high performance miniature two axis rate gyro which together with an electronic module provides two d.c. outputs corresponding to rate inputs about two orthogonal axes and is made by BAC costs about £1000 and weighs 56 grams plus 36 grams for the electronics. Such a package would be ideal for stabilisation purposes in certain homing heads.

5.5 ACCELEROMETERS

There are three types of linear accelerometer in common use:

(a) Spring-mass accelerometers usually called seismic mass accelerometers

(b) Piezo-electric accelerometers

(c) Force-balance accelerometers

Type (b) exhibit an electric charge across two faces proportional to the impressed force and hence acceleration but need a special charge amplifier if low frequency accelerations are to be recorded. Type (c) is really a more accurate spring-mass accelerometer and is used when great accuracy is required. Type (a) is the type most often employed in tactical missiles and consists of a mass suspended in a case by a low hysteresis spring; fluid damping is generally used. The force and hence deflection of the spring must be proportional to the acceleration and the design is such that there is one sensitive axis only with very little cross-coupling, although there are some designs with a more complicated suspension and two pick-offs so that two

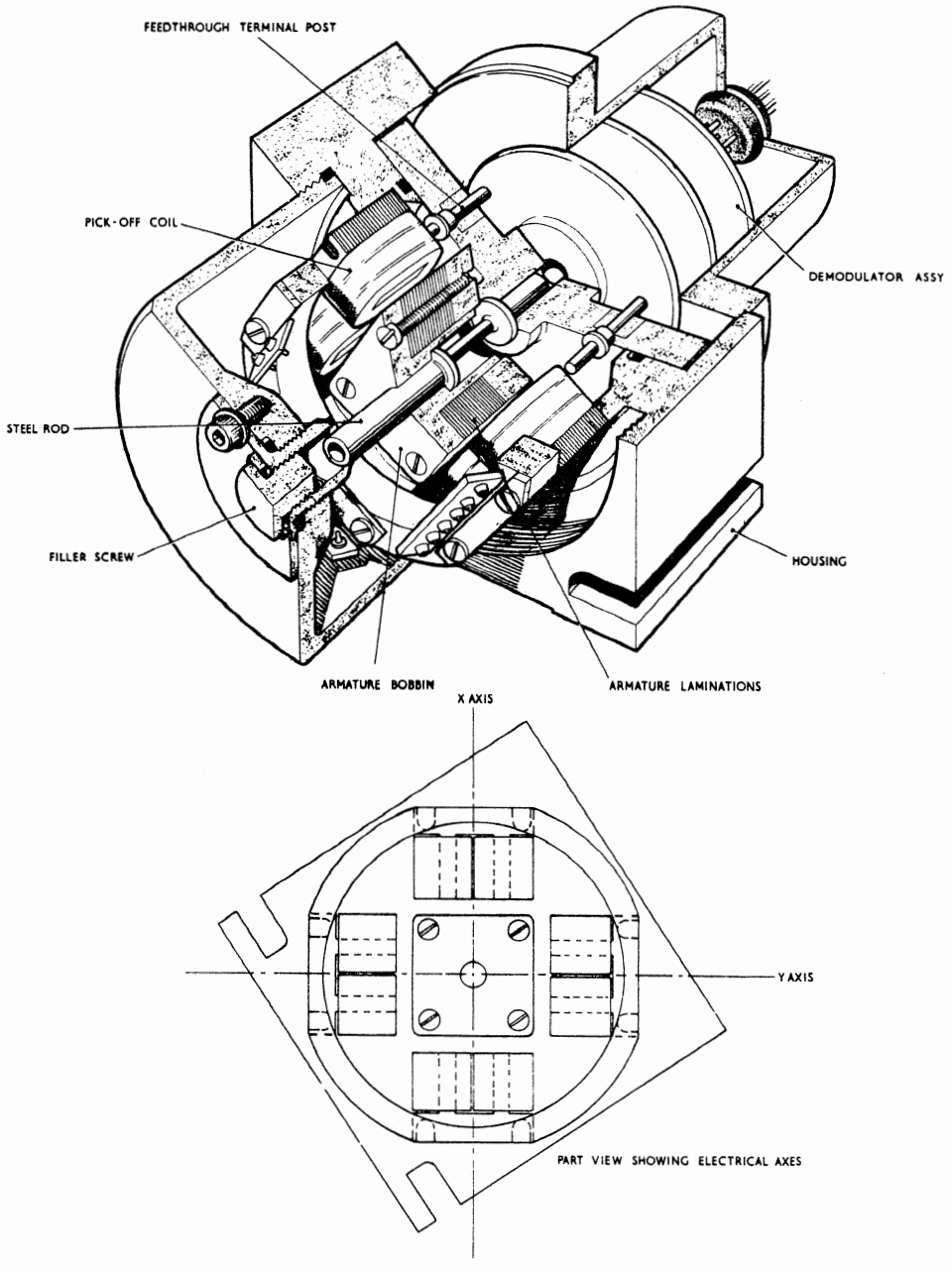


FIG. 5.5-1 Two axis accelerometer details

orthogonal accelerations can be measured. However, as the system is a conventional spring-mass arrangement there is a second order lag involved as with the rate gyro; a.c. pick-offs similar to those used with rate gyros measure displacement with respect to the case. Undamped natural frequencies in the range 80-100 Hz are typical. Again linearity and high resolution are comparatively easy to achieve, but to guarantee a bias much less than about ± 2% of full scale deflection under conditions of vibration, rough usage and wide ranges in temperature is not easy. Accelerometers for tactical missiles are likely to cost from £40-£200 depending on specification. Some two axis accelerometer details are shown in Fig 5.5-1.

5.6 RESOLVERS

In situations where the guidance is on the ground and the missile is not roll-position stabilised it is necessary to resolve the two guidance commands (i.e. "up-down" and "left-right") into freely rolling missile axes. Suppose the guidance demand up-down in V_1 and the missile was launched with the elevators horizontal but has since rolled anti-clockwise through an angle ϕ. If the elevator servos receive a command $V_1 \cos \phi$ and the rudder servos receive a command $-V_1 \sin \phi$, the missile will manoeuvre upwards always assuming that the aerodynamics are linear and the control surface movements are proportional to the input command. The induction resolver consists of a rotor and a stator, each with two windings whose electrical axes are at 90^0 to each other. The secondary voltages which result are proportional to the sine and cosine of the shaft angle. We therefore hold the rotor stationary in space by means of a roll gyro and allow the stator to rotate with the missile. The left-right guidance signal is applied to the other primary winding. The resulting relative voltages and forces from this arrangement are shown diagramatically in Fig 5.6-1. There is a coupling factor between primary and secondary windings but this has been ignored.

An alternative resolver using d.c. is shown in Fig 5.6-2. If the up-down guidance command is V_1 and the left-right command is V_2 then V_1, $V_2 - V_1$, $-V_2$ are applied to the points indicated. Two wipers at right angles will pick off linear functions of V_1 and V_2 which are well approximated to sines and cosines if the values of resistances are carefully chosen. Using eight tapping points as illustrated a reasonable approximation is obtained if the voltage at 45^0 is 0.75 times the voltage at 0^0. Fig 5.6-3 shows one approximation to a cosine for wiper angles from 0 to 360^0 for given values of

R_2/R_1 and R_3/R_1.

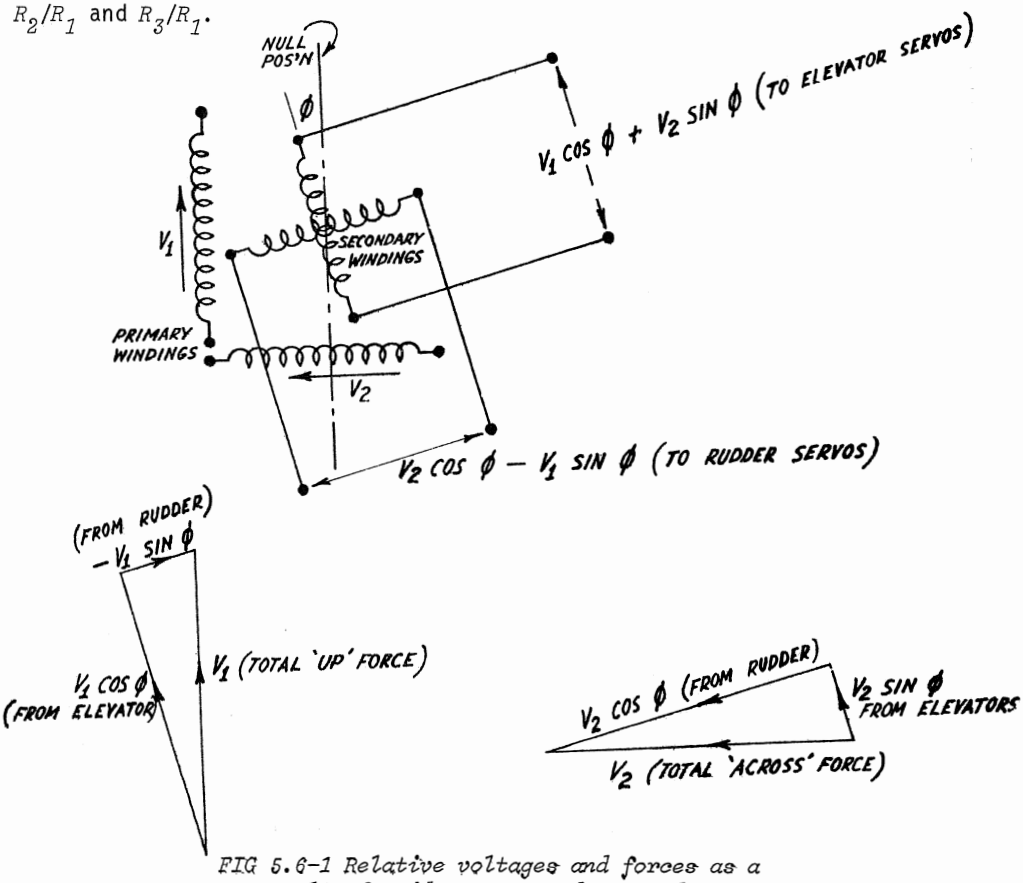


FIG 5.6-1 Relative voltages and forces as a result of guidance commands V_1 and V_2

5.7 ALTIMETERS

Radar altimeters are used in aircraft to indicate height above the ground or sea. A barometric altimeter indicates height above sea level or some other selected elevation. If a missile is required to fly at a given height above the ground for a distance of 20 or 30 km a simple barometric capsule or even a piezo electric pressure transducer should be accurate enough to indicate height providing this is not less than about 100 m. Below this height they are not suitable due to local small variations in atmospheric pressure and their limitation of discrimination and accuracy.

Radar altimeters are discussed by Hovanessian (2). Both FM/CW and pulsed systems can now operate down to altitudes of a metre or so but FM/CW systems appear to be more accurate in the range 0-10 m. Both types of altimeter can be made to indicate height continuously over a very wide range but this requires

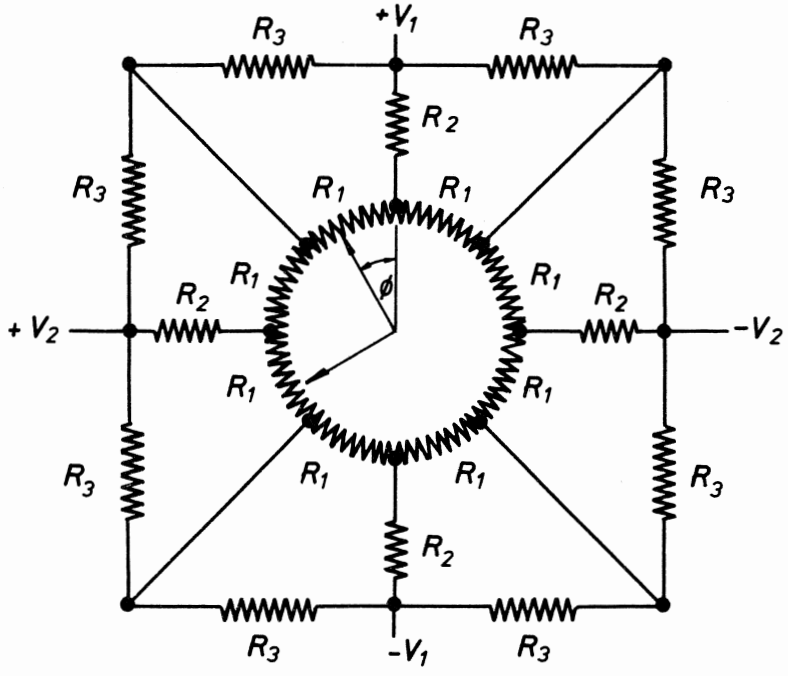


FIG 5.6-2 Ring resolver

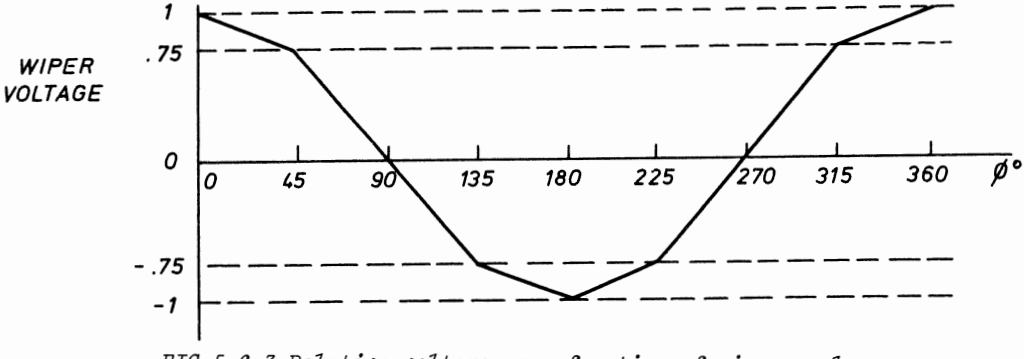


FIG 5.6-3 Relative voltage as a function of wiper angle

an elaborate design. If height measurement is required in the range 0-60 m only then a simpler instrument is possible weighing not more than about 2.5 kg. Both types can be designed with a large beamwidth permitting roll and pitch angles up to ± 25° or more. The distance measured tends to be the distance to the point of nearest return. A typical commercial FM/CW altimeter has an accuracy of ± 5% or ± 0.5 m below 10 m.

REFERENCES

1. SIMONS W.R. Errors in rate gyroscopes and some compensation techniques. Design Electronics, August 1966.

2. HOVANESSIAN S.A. Radar detection and tracking systems. Artec House Inc. 1973.

CHAPTER 6

AUTOPILOT DESIGN

6.1 INTRODUCTION

An autopilot is a closed loop system and it is a minor loop inside the main guidance loop; not all missile systems require an autopilot. A missile will manoeuvre up-down or left-right in an apparently satisfactory manner if a control surface is moved or the direction of thrust altered. If the missile carries accelerometers and/or gyros to provide additional feedback into the missile servos to modify the missile motion then the missile control system is usually called an autopilot, but this definition is not universally accepted. Broadly speaking autopilots either control the motion in the pitch and yaw planes, in which case they are called lateral autopilots, or they control the motion about the fore and aft axis in which case they are called roll autopilots. For a symmetrical cruciform missile pitch and yaw autopilots are identical; one injects a g bias in the vertical plane to offset the effect of gravity but this does not affect the design of the autopilot. Roll autopilots serve quite a different purpose and will be considered separately.

The requirements for a good lateral autopilot are very nearly the same for command and homing systems but it is more helpful initially to consider those associated with command systems where the guidance receiver produces signals proportional to the misalignment of the missile from the line of sight (LOS). A simplified closed-loop block diagram for a vertical or horizontal plane guidance loop without an autopilot is shown in Fig 6.1-1.

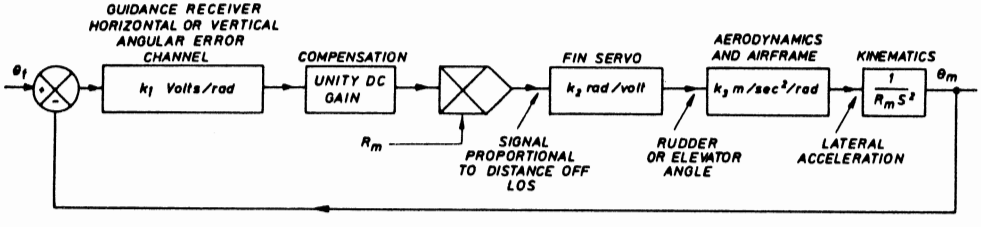


FIG 6.1-1 A vertical or horizontal plane guidance loop

The target direction θ_t is determined by the target tracker (we shall put the target and missile tracking systems together in the next chapter). Let the guidance receiver gain be k_1 volts/rad (misalignment). The guidance signals are then invariably phase advanced to ensure closed loop stability. In order to maintain constant sensitivity to missile linear displacement from the LOS, the signals are also multiplied by the measured or assumed missile range R_m before being passed to the missile servos. This means that the effective d.c. gain of the guidance error detector is k_1 volts/m. If the missile servo gain is k_2 rad/volt and the control surfaces and airframe produce a steady state lateral acceleration of k_3 m/sec^2/rad then the guidance loop has a steady state open loop gain of $k_1\ k_2\ k_3$ m/sec^2/m or $k_1\ k_2\ k_3$ sec^{-2}. The loop is closed by two inherent integrations from lateral acceleration to lateral position. Since the error angle is always very small one can say that the change in angle is this lateral displacement divided by the instantaneous missile range R_m. Hence the x R_m and ÷ R_m factors can be regarded as cancelling each other out. It is argued later that this assumption of a double integration is not strictly correct but that it is a reasonable approximation for guidance loop stability calculations. Consider now the possible variation in the value of the aerodynamic gain k_3. If, due to uncertainties and changes in the c.g. because of

(a) propellant consumption

and (b) manufacturing tolerances

and changes in c.p. due to

(a) changes in incidence

(b) changes in missile speed

(c) manufacturing tolerances

the value of k_3 could easily change by a factor of 5 to 1 if the static margin should change from say 2 cm to 10 cm in a missile about 2 m long. If, in addition, there can be large variations in the dynamic pressure $\frac{1}{2}\ \rho U^2$ due to changes in height and speed (many missiles have a boost-coast velocity profile), then the overall variation in aerodynamic gain could easily exceed 100 to 1. Also, we have already seen that the weathercock frequency of the missile which is a measure of the speed of response of the airframe in producing a lateral acceleration is a function of the static margin and the aerodynamic derivatives. And finally, since the missile response depends on the semi non-dimensional form of the aerodynamic derivatives, it follows that all aspects of the missile response will vary as the mass and inertia

vary with propellant usage. A satisfactory guidance loop cannot be engineered if very large tolerances exist inside the loop.

Since the bandwidth and damping of the guidance loop are optimised to produce minimum miss distance in the presence of target motion and noise of various sorts it follows that the receiver designer as well as the missile control system designer have to work to fairly close tolerances on gain, and in the latter case on dynamic response as well. No problem exists in maintaining a constant missile servo gain. If now it is known that the c.g. position will not vary from round to round and that it will not shift during flight, that the c.p. position will not change either, that the air density and missile velocity will remain constant and that the aerodynamics are perfectly linear (i.e. normal force is proportional to incidence), then he may have no worries. There are some airframe configurations (usually with wings with long root chords) where the c.p. changes very little with changes in Mach number and incidence. Moreover, if one can design a missile for high stability and hence low manoeuvrability (i.e. with a static margin of 5% or more of the length) then a small shift in c.p. or c.g. should not result in a large percentage change in static margin. Many anti-tank missiles are in this category. The maximum g requirements for such missiles which are designed to hit relatively slow moving targets are typically 3 to 4 g at the most. In addition, they are usually boosted and sustained to fly at a near-constant velocity. In such a case the problem of keeping the aerodynamic gain roughly constant hardly arises.

We are now in a position to discuss the main objectives in lateral autopilot design.

6.2 LATERAL AUTOPILOT DESIGN OBJECTIVES

The maintenance of near-constant steady state aerodynamic gain

Most guidance loops are conditionally stable i.e. a variation in gain either up or down will erode stability margins. The missile control designer might reasonably be asked to keep to a tolerance of ± 25% of a nominal value. However many command systems use "feedforward terms" to improve accuracy and these will be discussed in the next chapter. To make full use of these the control system should have a close tolerance on steady state gain, ideally ± 10% of a nominal value or better. In homing systems, we shall see that the performance is seriously degraded if the "kinematic gain" varies by more than about ± 30% of an ideal value. Since the kinematic gain depends on the control system gain, the homing head gain and the missile-target relative

velocity, and the latter may not be known very accurately, it is not expected that the missile control designer will be allowed a tolerance of more than ± 20%.

So we arrive at a general conclusion that an open loop missile control system is probably not acceptable for highly manoeuvrable missiles which have very small static margins especially those which do not operate at a constant height and speed.

To increase the weathercock frequency

A high weathercock frequency is essential for the stability of the guidance loop. Since the rest of the loop consists essentially of two integrations and a d.c. gain it follows that if there are no dynamic lags in the loop whatsoever we have 180^0 phase lag at all frequencies open loop. To obtain stability we can pass the guidance error signal through a phase advance network whose transfer function is

$$\frac{1 + Ts}{1 + \alpha Ts}, \text{ where } \alpha < 1$$

In the frequency domain this approximates to $1/\alpha$ for $\omega \alpha T > 1$ and hence this operation of adding the approximate differential to the error signal has the inevitable effect of seriously degrading the signal-to-noise ratio as some high frequency noise will almost certainly be present. The maximum amount of phase advance ϕ_{max} occurs when

$$\omega T = 1/\sqrt{\alpha}$$

giving

$$\tan \phi_{max} = (1 - \alpha)/2\sqrt{\alpha} \qquad (6.2\text{-}1)$$

For

$\alpha = 0.3$ $\quad \phi_{max} = 31^0$

$\alpha = 0.2$ $\quad \phi_{max} = 42^0$

$\alpha = 0.1$ $\quad \phi_{max} = 55^0$

If $\alpha = 0.2$, the network is often referred to as a "5 to 1 phase advance network" and if $\alpha = 0.1$ as a "10 to 1 phase advance network". The point that is being stressed at the moment is that amplitudes multiply and phases add. We shall point out in the next chapter that one can do a little better than this with a more complex type of phase advance transfer function but the fundamental problem still remains: if one requires more than about 60^0 phase advance one has to use several networks in series and the deterioration in signal-to-noise ratio tends to be catastrophic. Hence, although the number is somewhat arbitrary, designers do tend to limit the amount of phase advance that is going to be required to about 60^0. This means that if one is

going to design a guidance loop with a minimum of 45^{o} phase margin, the total phase lag permissible from the missile servo and the aerodynamics at guidance loop unity gain cross-over frequency will be 15^{o}. Hence the servo must be relatively very fast and the weathercock frequency should be much faster than guidance loop undamped natural frequency - i.e. the open loop unity gain cross-over frequency. This may not be a practical proposition for an open loop missile control system especially at the lower end of the missile speed range and with a small static margin.

To increase the weathercock damping

We have seen that the weathercock mode is very underdamped, especially with a large static margin and at high altitudes. This has several undesirable effects. Firstly, a badly damped oscillatory mode results in a large r.m.s. output to broad band noise. The r.m.s. incidence is unnecessarily large and this results in a significant reduction in range due to the induced drag. The accuracy of the missile will also be degraded somewhat. A sudden increase in signal which could occur after a temporary loss of signal will result in a large overshoot both in incidence and in achieved lateral g. This might cause stalling and in any case it would mean that the airframe would have to be stressed to stand nearly twice the maximum designed steady state g.

To reduce cross-coupling between pitch and yaw motion

We have discussed this aspect of Euler's equations, see equation 4.2-7. It is clear that if synthetic damping is introduced in both the lateral modes then the mean pitch and yaw rates will be reduced. This must reduce cross-coupling. Also if roll rates can be kept low the effects of cross-coupling should be negligible.

To assist in gathering

In a command system the missile is usually launched some distance off the line of sight and toe-in and superelevation are added such that if there are no disturbances, the missile will fly into the beam or LOS. Thrust misalignment, biases and cross winds all contribute to dispersion. A closed loop missile control system (i.e. an autopilot), if it behaves like any other closed loop system should be reasonably resistant to outside disturbances. To improve guidance accuracy the systems engineer will want the narrowest guidance beam possible. The majority of supersonic CLOS and beam riders would stand a very small chance of gathering if they possessed no autopilot. Before proceeding to a detailed discussion of some particular autopilots it may be helpful to classify them as shown in Fig 6.2-1.

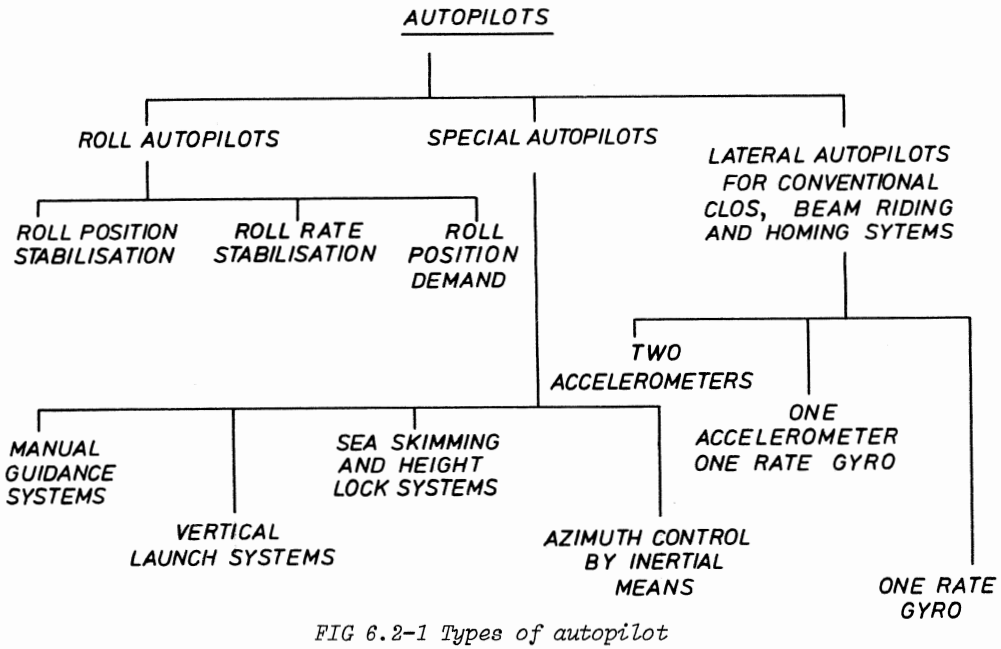


FIG 6.2-1 Types of autopilot

6.3 A LATERAL AUTOPILOT USING ONE ACCELEROMETER AND ONE RATE GYRO

An arrangement whereby an accelerometer provides the main feedback and a rate gyro is used to act as a damper is common in many high performance command and homing missiles. Identical autopilots are used to control the pitch and yaw motions if the missile has two planes of symmetry, so we need consider one channel only, the yaw autopilot say. The accelerometer is placed well forward of the c.g. probably about a half to two thirds of the distance from the c.g. to the nose. Its sensitive axis is in the direction oy. One has to avoid placing the accelerometer at an antinode of the principle bending mode of the missile otherwise the vibration pick-up at this point may result in destruction of the missile. If the missile servo can respond to this body bending frequency the resulting fin movement will tend to reinforce this natural mode. The rate gyro is ideally placed at a node where the angular movement due to vibration is a minimum. The rate gyro's sensitive axis is along oz i.e. its output is proportional to yaw rate r.

Fig 6.3-1 shows the arrangement in transfer function form where

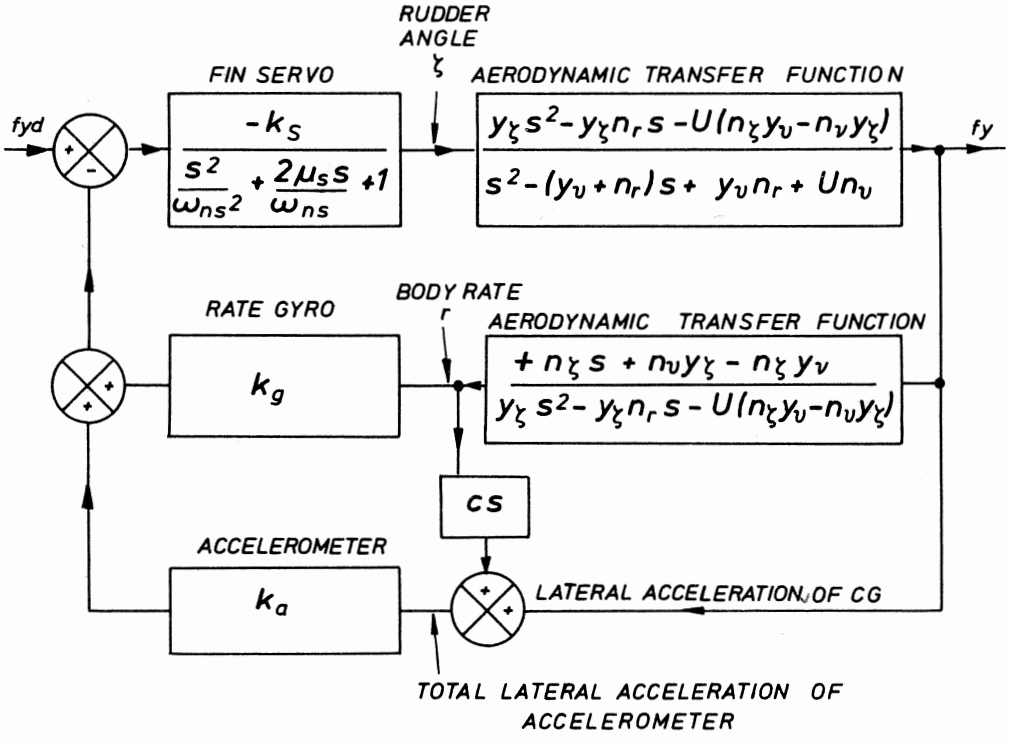


FIG 6.3-1 Lateral autopilot with one accelerometer and one rate gyro

f_{yd} and f_y are the demanded and achieved lateral accelerations of the c.g.

$-k_s$ is the servo gain in radians rudder/volt

k_a is the accelerometer gain in volts/m/sec^2

k_g is the rate gyro gain in volts/rad/sec

and c is the distance of the accelerometer sensitive axis forward of the c.g.

The servo gain has to be negative since the transfer function f_y/ζ has a negative gain. It is assumed that the fin servo dynamics are adequately defined by a second order lag. The two aerodynamic transfer functions when multiplied together give the transfer function r/ζ, see equation 4.6-8. Manipulation yields the following autopilot closed loop transfer function

$$\frac{f_y}{f_{yd}} = \frac{-y_\zeta s^2 + y_\zeta n_r s + U\,(n_\zeta y_v - n_v y_\zeta)}{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_o} \qquad (6.3\text{-}1)$$

where $a_o = \frac{1}{k_s}(y_v n_r + U n_v) + k_a U (n_\zeta y_v - n_v y_\zeta) + k_g (n_\zeta y_v - n_v y_\zeta)$

$$a_1 = \frac{1}{k_s}\{\frac{2\mu_s}{\omega_{ns}}(y_v n_r + U n_v) - (y_v + n_r)\} - k_g n_\zeta + k_a \{y_\zeta n_r + c\,(n_\zeta y_v - n_v y_\zeta)\}$$

$$a_2 = \frac{1}{k_s}\{1 + \frac{y_v n_r + U n_v}{\omega_{ns}^2} - 2\mu_s \frac{(y_v + n_r)}{\omega_{ns}}\} + k_a (-y_\zeta - c\, n_\zeta)$$

$$a_3 = \frac{1}{k_s \omega_{ns}}\{2\mu_s - \frac{(y_v + n_r)}{\omega_{ns}}\}$$

$$a_4 = \frac{1}{k_s \omega_{ns}^2}$$

In order to design the autopilot for a rear controlled missile we must make a first estimate of the required aerodynamic derivatives. We calculate that the missile may have to develop a lateral acceleration of 250 m/sec^2 at a forward velocity of 500 m/sec (lateral g requirements are discussed in the next chapter). The missile is to be of the boost coast type and the maximum velocity will be $\sqrt{2}$ x 500 m/sec and the minimum velocity will be $500/\sqrt{2}$ m/sec i.e. a speed variation of 2 to 1 overall. We make the assumption that normal force will be proportional to incidence and we limit body incidence to 0.2 radians. Hence

$$y_\beta \beta = 250 = y_\beta \times 0.2 \therefore y_\beta = 1250$$

But $\beta = v/U$ for small angles $\therefore y_v = 1250/500 = 2.5$. However since some of the lift of body + wings + control surfaces is lost with a tail controlled missile we should allow at least + 10% on this estimate, say $y_v = 3.0$. We estimate the missile length to be 2m, the all-burnt mass to be 52 kg and the all-burnt lateral moment of inertia C to be 14 kg m^2. The value of n_v will depend on y_v and the size of the static margin which we have some control over since we can dictate to some extent the position of the wings. Let us go initially to a fairly healthy mean static margin of 4-5% of the total length bearing in mind that this will vary somewhat with incidence and Mach number.

Now $N_v = Y_v x^*$ where x^* is the static margin and $n_v = N_v/C$ and $y_v = Y_v/m$ If x^* is 0.09m, i.e. 4.5% of the missile length then we find that $n_v = 1.0$. We now consider the strength of rudder required. If the rudder moment arm is three quarters of half the missile length i.e. 0.75m and the static margin is 0.09m then to balance the stability moment with a rudder deflection of 0.2 radians we need $y_\zeta = 1250 \times 0.09/0.75 = 150$. However a steady lateral

acceleration results in a steady body rate and hence an opposing damping moment which also opposes the rudder moment. Suggest y_ζ = 180. The damping derivative n_r is not necessary to specify so we take a figure based on similar designs i.e. n_r = -3. Since we know the control surface moment arm we find n_ζ = 500. Tentatively we put the accelerometer 0.5m forward of the c.g. Summarising therefore (and putting in the correct algebraic signs) our control parameters are

$$U = 500 \qquad y_\zeta = +180$$
$$y_v = -3.0 \qquad n_\zeta = -500$$
$$n_v = +1.0 \qquad c = 0.5$$
$$n_r = -3.0$$

This leaves us with 5 parameters over which we have some control and which we wish to optimise viz k_s, ω_{ns}, μ_s, k_a, k_g; we shall attempt in the first place, to design the autopilot without any compensation circuits.

What will the specification look like? Suppose the nominal guidance loop gain is 10. With phase advance the frequency at which open loop gain cross-over occurs will be increased from $\sqrt{10}$ to about 5 rad/sec. So we are looking for a small phase lag from the autopilot in the critical frequency range say 2-12 rad/sec. A requirement for a maximum autopilot phase lag of 15^0 at 5 rad/sec and 30^0 at 10 rad/sec would be just adequate. The system uses feed-forward terms so we would expect to keep to a given closed loop gain ± 10% but we might be allowed a few per cent more than this at the extremes of missile speed. We would expect to see something in the specification to control the degree of stability, such as percentage overshoot in the time domain, and gain and phase margin in the frequency domain. In addition there may be a limit imposed on the closed loop resonant gain in order to limit the equivalent noise bandwidth. The autopilot will also have to be satisfactory from the gathering point of view and a very close watch will have to be kept on the servo gain and bandwidth to avoid fin rate saturation in the presence of noise. In this connection the designer will try for the lowest possible servo gain and the lowest possible servo bandwidth as these measures will result in a reduction of r.m.s. fin rate.

The reader will already know that a fundamental quantity in any closed loop is the open loop gain, and this is the product of the feedforward gain and the feedback gain. So for convenience we can design the system as a nominal 1 to 1 system i.e. the closed loop gain is unity. Since the autopilot is a type 0 feedback system and the rate gyro provides some of the feedback

we can without loss of generality put the accelerometer gain to 0.8 as a first try. If in practice it is more convenient to put the accelerometer gain to a tenth of this value then the open loop gain is easily restored to its correct value by increasing the servo gain by a factor of 10. Let us set the servo undamped natural frequency ω_{ns} to 180 rad/sec, a figure which can be obtained with a good hot gas or a conventional pneumatic servo and $\mu_s = 0.5$. From a first try on an analogue simulator it looks as though $k_s = 0.007$ rad/volt and $k_g = 30$ volts/rad/sec will result in a reasonable autopilot.

We now insert these values in equation (6.3-1) and find that

$$a_o = 6.39 \times 10^5 \simeq k_a U n_\zeta y_v$$
$$a_1 = 1.636 \times 10^4 \simeq - k_g n_\zeta$$
$$a_2 = 2.06 \times 10^2 \simeq 1/k_s + k_a (- y_\zeta - c n_\zeta)$$

This coefficient is approximately zero if $c = 0$ which shows how vital it is to place the accelerometer forward of the c.g.

$$a_3 = 0.82 \simeq 2\mu_s/k_s \omega_{ns}$$

$$a_4 = 4.41 \times 10^{-3} = 1/k_s \omega_{ns}^2$$

In order to consider the stability and modes of oscillation of this fourth order system we need to consider the coefficients of the denominator only of the closed loop transfer function. The characteristic equation is obtained by rewriting the transfer function as a differential equation and equating the left hand side to zero thus:

$$\left(\frac{0.689s^4}{10^8} + \frac{1.282s^3}{10^6} + \frac{3.222s^2}{10^4} + \frac{2.587s}{10^2} + 1\right) 6.39 \times 10^5 f_y = 0 \quad (6.3\text{-}2)$$

Factorisation of this equation by trial and error is not possible since it will almost certainly factorise into two quadratics each with complex roots. The standard method of solving such an equation is to write it in this form

$$s^4 + 2as^3 + bs^2 + 2cs + d = 0$$

and to solve the cubic $2k^3 - bk^2 + 2(ac - d)k - a^2d + bd - c^2 = 0$

From this cubic equation one real value of k can always be found. The values of e and f are then obtained from the equations

$$a^2 + 2k = b + e^2 \text{ and } k^2 = d + f^2$$

whence the values of s can be obtained from the two quadratics

$$s^2 + (a - e)s + (k - f) = 0$$
$$\text{and } s^2 + (a + e)s + (k + f) = 0$$

Equation (6.3-2) thus factorises to:

$$\left(\frac{s^2}{65.3^2} + \frac{2 \times 0.611s}{65.3} + 1\right)\left(\frac{s^2}{185^2} + \frac{2 \times 0.286s}{185} + 1\right) = 0 \qquad (6.3\text{-}3)$$

It is easily shown by the residue method that the response to a step input is dominated by the slower mode defined by ω_n = 65.3 rad/sec and μ = 0.611; in fact the amplitude of this slower mode is about seven times larger than the faster mode. This main mode is well damped. The fact that the faster mode is rather poorly damped is not serious and it contributes relatively little to the system response. The open loop weathercock mode is defined by $\omega_n^2 = Un_v + y_v n_r$ = 509, and $2\mu\omega_n = y_v + n_r$. Hence ω_n = 22.5 and μ = 0.13. Thus we have increased the speed of response of this slow mode by a factor of 3 to 1 and we have increased the damping to a very respectable value. What aids are there available to the designer in arriving at a satisfactory fourth-order response?

A much clearer insight into the behaviour of linear systems is obtained by using a normalisation procedure. Consider the transfer function

$$\frac{\theta_o}{\theta_1} = \frac{1}{a_2 s^2 + a_1 s + 1}$$

We are familiar with the method of writing it thus

$$\frac{\theta_o}{\theta_1} = \frac{1}{s^2/\omega_n^2 + 2\mu s/\omega_n + 1} \text{ where } \omega_n = \frac{1}{\sqrt{a_2}}$$

and we know that the time or frequency response can be normalised against non-dimensional time $\omega_n t$ or non-dimensional frequency ω/ω_n. In other words we have reduced the number of variables by one and the type of response depends only on the normalised value of a_1. We also note that this value 2μ is likely to be a small number say between 1 and 2. Graham and Lathrop (1) have shown that systems of any order with identical normalised transfer functions possess identical pole-zero geometry, differing only in magnitude proportional to the normalisation frequency ω_o.

Thus for a fourth-order system we write $\omega_o = (1/a_4)^{\frac{1}{4}}$ and $\omega_o . S = s$ such that our characteristic equation can be rearranged from equation (6.3-2) to read

$$S^4 + 1.69S^3 + 3.88S^2 + 2.81S + 1 = 0 \qquad (6.3\text{-}4)$$

where ω_o = 109.7

We have already considered the "optimum" damping ratio for a second order system i.e. we have to control the coefficient of S only. In a fourth-order system the problem is the same in principle except now we have to control

the coefficients of s^1, s^2 and s^3! The optimum value of these coefficients will clearly depend on the performance index selected. There is general agreement among control engineers that for any order system the transfer function which minimises the integral of time x absolute error ITAE to a step input will have a good transient response and a good frequency response. For systems with no closed loop zeroes the optimum transfer function coefficients are shown in Table 6.3-1.

TABLE 6.3-1 DENOMINATOR OF NORMALISED ITAE OPTIMUM SYSTEMS WITH NO CLOSED LOOP ZEROES

Order	Denominator
2	$s^2 + 1.4s + 1$
3	$s^3 + 1.75s^2 + 2.15s + 1$
4	$s^4 + 2.1s^3 + 3.4s^2 + 2.75s + 1$

The authors consider that the method of designing third and fourth order systems with the aid of these normalised forms is particularly useful in this context. The root locus method gives valuable insight when one is varying one parameter at a time, such as the loop gain. In the case of a lateral autopilot, a change in aerodynamic gain due to a change in missile speed say results simultaneously in changes in all the other aerodynamic parameters such as the weathercock frequency and the weathercock damping. We note that our normalised denominator as given in equation (6.3-4) is very similar to the I.T.A.E. optimum form. Now a little experience with such systems leads one to the conclusion that there is a fair amount of latitude about this so called optimum; the system response is not very sensitive to small changes in the coefficients about these optimum values. The coefficient of s^3 is about 80% of the optimum value. This suggests that one of the modes may be rather underdamped. If we look at the terms which comprise a_3 we see that it is proportional to the servo damping ratio to a good first approximation. Those readers who have worked with the root locus technique will realise that the simplest way to improve the damping of a closed loop mode is to start with an open loop mode which is well damped. So a fundamental and important principle emerges: if the servo response is well damped (μ = 0.7 to 1.0 say) the design of a well damped autopilot is greatly facilitated.

Finally, before considering how the autopilot response changes when the missile speed changes it is instructive to look at the autopilot transfer function using the Routh Hurwitz stability criterion. The conditions for stability are

$$a_1 (a_3 a_2 - a_1 a_4) > a_0 a_3^2 \qquad (6.3\text{-}5)$$

It does not matter whether we use the original transfer function or the normalised one, we will arrive at the same answer. In the latter case $a_4 = a_o = 1$ and hence (6.3-5) reduces to

$$a_1 (a_3 a_2 - a_1) > a_3^2 \qquad (6.3\text{-}6)$$

Towill (2) has shown that the absolute minimum normalised value of $a_2 = 2$ when $a_1 = a_3$. Examination of the terms which make up a_2 will convince us that we should try to place the accelerometer as far forward as possible since a_2 increases as c increases. Indeed, equation (6.3-6) is saying that there are two conditions for stability, the one explicitly stated and another i.e.

$$a_3 a_2 - a_1 > 0$$

One sees immediately that an increase in the value of a_2 must assist in promoting stability. This is just not so for the other coefficients a_3 and a_1. Now the value of a_1 is largely dependant on the rate gyro gain k_g. It is easy to show on an analogue simulator that too large a value of rate gyro gain causes instability. Similarly if the servo damping is really excessive (increasing the value of a_3) equation (6.3-6) is warning us that instability can arise due to this also.

We will now see how the autopilot response varies as the missile speed changes. Investigation of wind tunnel results of some typical rather slender low aspect ratio supersonic missile configurations shows that to a good first approximation the normal force for a given incidence is proportional to $\frac{1}{2}\rho U^2$ and is only weakly dependant on Mach number in the speed range under consideration. If therefore the speed increases by a factor of $\sqrt{2}$ we will assume that y_ζ, n_ζ and n_r increase by a factor of 2 and that y_v and n_v increase by a factor of $\sqrt{2}$, see chapter 4.7.

Inserting these new values in equation (6.3-1) yields the denominator coefficients shown in Table 6.3-2.

TABLE 6.3-2 DENOMINATOR COEFFICIENTS OF AUTOPILOT TRANSFER FUNCTION

	$U = 500/\sqrt{2}$	$U = 500$	$U = 500 \times \sqrt{2}$
a_o	1.821×10^5	6.39×10^5	2.37×10^6
a_1	8.291×10^3	1.636×10^4	3.204×10^4
a_2	1.746×10^2	2.06×10^2	2.675×10^2
a_3	0.806	0.82	0.839
a_4	4.41×10^{-3}	4.41×10^{-3}	4.41×10^{-3}

For $U = 500/\sqrt{2}$ the closed loop transfer function reduces to:

$$\frac{f_y}{f_{yd}} = \frac{0.907\left(-\frac{5.45s^2}{10^4} - \frac{8.41s}{10^4} + 1\right)}{\frac{2.422s^4}{10^8} + \frac{4.43s^3}{10^6} + \frac{9.59s^2}{10^4} + \frac{4.55s}{10^2} + 1}$$

$$= \frac{0.907\ (-\ 0.0238s + 1)(.0229s + 1)}{\left(\frac{s^2}{37.1^2} + \frac{2 \times 0.739s}{37.1} + 1\right)\left(\frac{s^2}{176.6^2} + \frac{2 \times 0.362s}{176.6} + 1\right)}$$

The normalised denominator reads $S^4 + 2.280S^3 + 6.161S^2 + 3.650S + 1$ and ω_o = 80.2 rad/sec.

For $U = 500$ the closed loop transfer function as we have seen reduces to

$$\frac{f_y}{f_{yd}} = \frac{1.03\left(-\frac{2.73s^2}{10^4} - \frac{8.18s}{10^4} + 1\right)}{\frac{0.689s^4}{10^5} + \frac{1.282s^2}{10^6} + \frac{3.222s^2}{10^4} + \frac{2.587s}{10^2} + 1}$$

$$= \frac{1.03\ (-\ 0.0169s + 1)(0.0161s + 1)}{\left(\frac{s^2}{65.3^2} + \frac{2 \times 0.611s}{65.3} + 1\right)\left(\frac{s^2}{185^2} + \frac{2 \times 0.286s}{185} + 1\right)}$$

The normalised denominator reads

$$S^4 + 1.69S^3 + 3.88S^2 + 2.81S + 1$$

and ω_o = 109.7 rad/sec

For $U = 500 \times \sqrt{2}$ the closed loop transfer function reduces to

$$\frac{f_y}{f_{yd}} = \frac{1.11\left(-\frac{1.364s^2}{10^4} - \frac{8.18s}{10^4} + 1\right)}{\frac{1.861s^4}{10^8} + \frac{1.129s^3}{10^6} + \frac{1.129s^2}{10^2} + \frac{1.352s}{10^2} + 1}$$

$$= \frac{1.11\ (-\ 0.01211s + 1)(0.01129s + 1)}{\left(\frac{s^2}{111.7^2} + \frac{2 \times 0.142s}{111.7} + 1\right)\left(\frac{s^2}{785^2} + \frac{2 \times 0.202s}{785} + 1\right)}$$

The normalised denominator reads $S^4 + 1.250S^3 + 2.617S^2 + 2.058S + 1$ and ω_o = 152.2 rad/sec.

The closed loop frequency responses for these three missile velocities are shown in Fig 6.3-2 and the responses to an input step demand of 50 m/sec^2 are shown in Fig 6.3-3. We are now in a position to say whether we are likely to meet the specification.

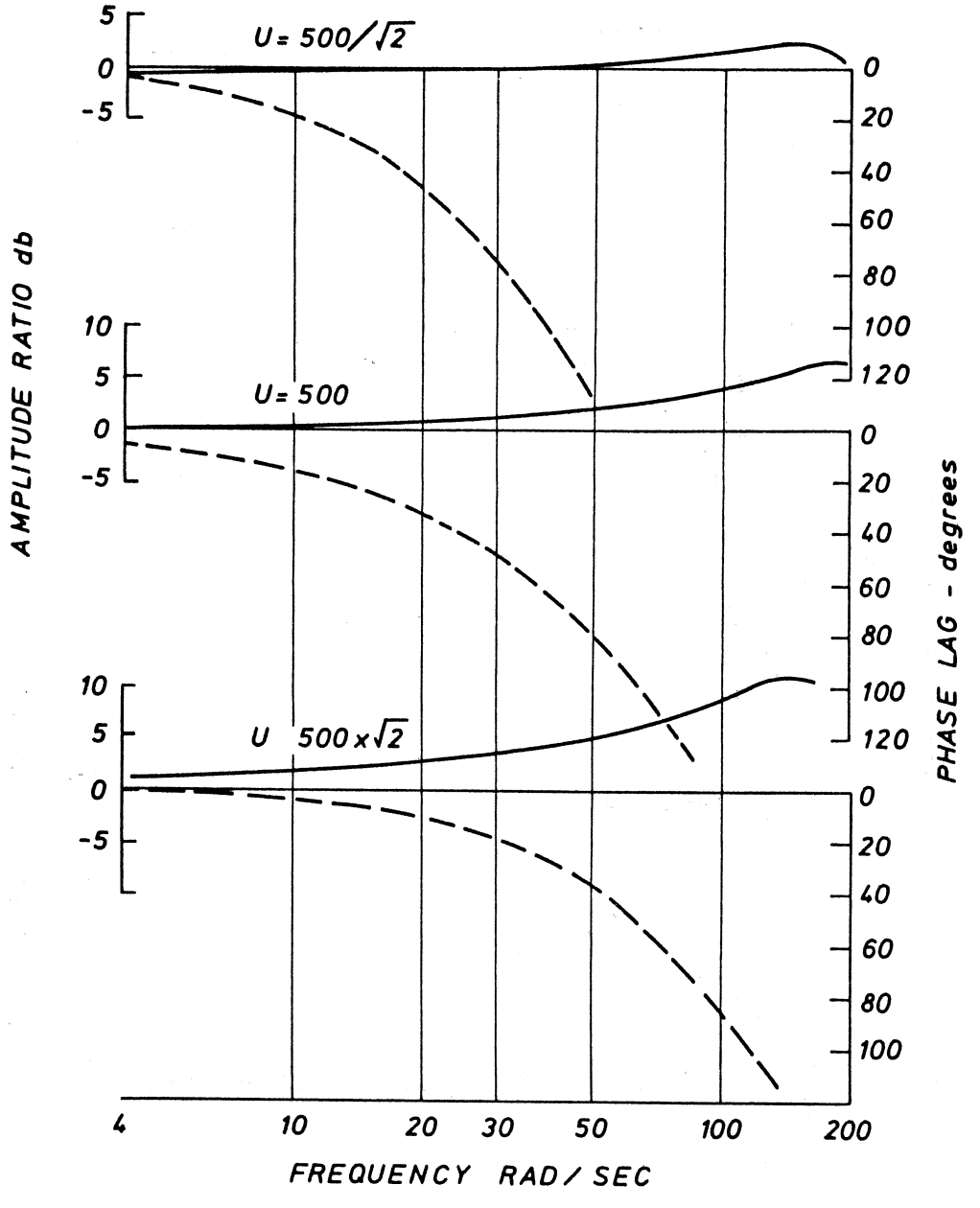


FIG 6.3-2 Lateral autopilot closed loop frequency responses

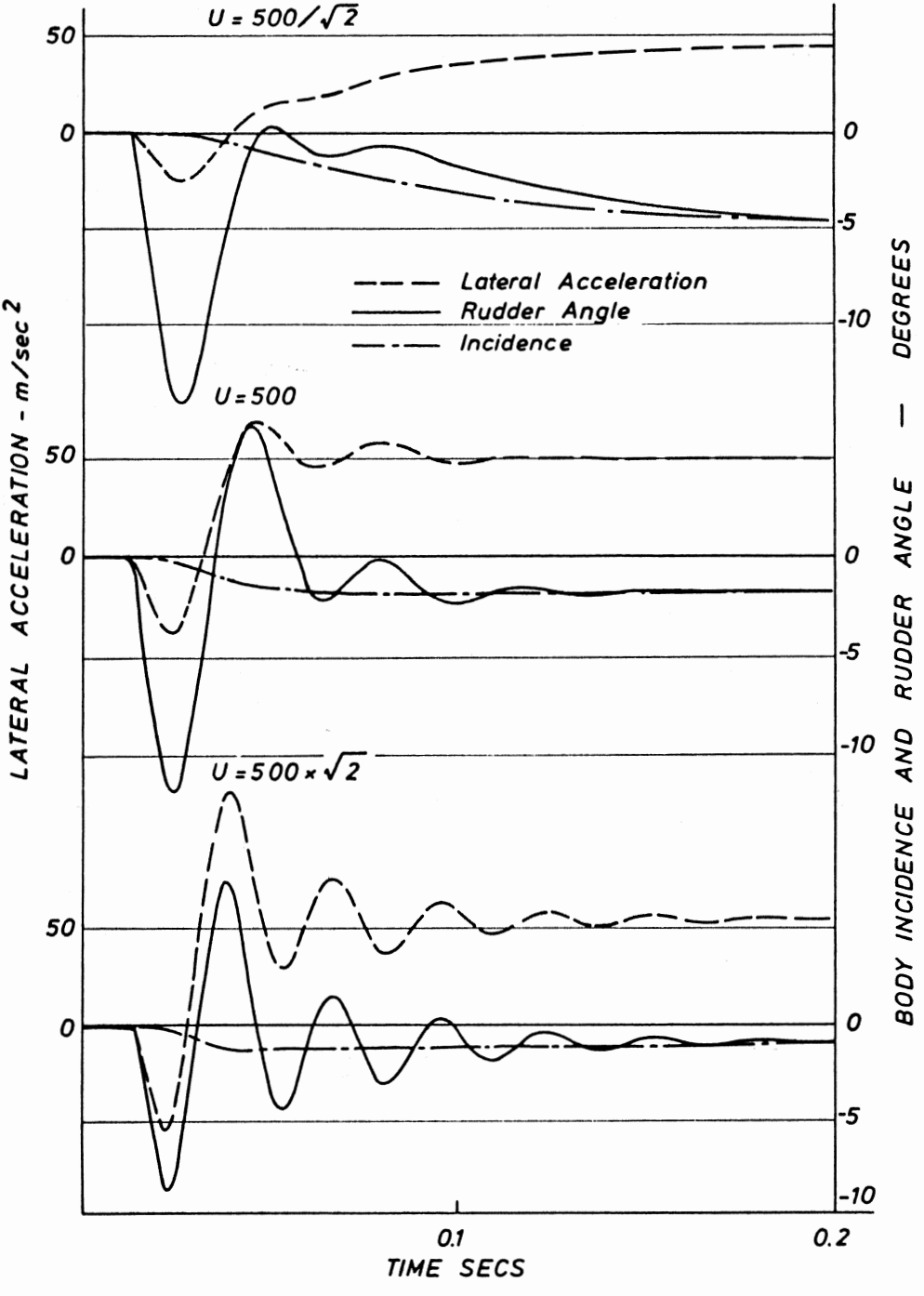


FIG 6.3-3 Lateral autopilot response to an input step demand for 50 m/sec²

(a) The closed loop gain varies from 0.907 to 1.11 a variation of -9.3% to +11% of the nominal value of unity. We originally had in mind a figure of ± 10%. This variation is due to a change in aerodynamic gain of 4 to 1 very nearly.

(b) The autopilot easily meets the requirement concerning phase lag in the specified frequency range. The reason why the phase lag is small at these low frequencies is that the undamped natural frequency of the dominant mode has been increased. Table 6.3-3 summarises the position.

TABLE 6.3-3 COMPARISON BETWEEN OPEN AND CLOSED LOOP MODES

Missile speed m/sec	Open loop		Closed loop	
	Undamped weathercock frequency rad/sec	Damping ratio	Dominant mode undamped frequency rad/sec	Damping ratio
$500/\sqrt{2}$	15.9	0.10	37.1	0.739
500	22.5	0.133	65.3	0.611
$500 \times \sqrt{2}$	32.0	0.16	111.7	0.142

The reason for this increased speed of response closed loop is easily seen from Fig 6.3-3. The rudder is allowed to go over a long way initially because the feedback is so small. The strongly phased advanced feedback however soon reverses the rudder in an attempt to prevent overshooting. This closed loop system therefore is fundamentally no different in action from any other feedback system. An increase in the speed of response depends not only on the transfer function itself but also on the physical ability to meet the power, rate, angle etc requirement implicit in this transfer function. Ignoring stability considerations at the moment, we could apparently, increase the speed of response still further by doubling the servo gain. As a result of a given input the rudder will now try, initially, to deflect twice as far as before. Since the steady state lateral acceleration will be approximately unaltered by this change in feedforward gain it follows that the response should be faster. But there is clearly a limit to the useful deflection we can use and in any case the servo will have to develop twice the angular rate as before; and this may be beyond its capability. Therefore, if we are to believe in the transfer functions we write down we have to limit some of the gains in a system or sub-system.

(c) We also note that the closed loop system, in general, is well damped, not because we have increased the viscosity of the air, but because we have regulated the timing or phasing of the torque producing mechanism by suitably

shaping the feedback. Nevertheless in the high speed range it is seen that both modes are under-damped. The dominant mode damping is greatly improved if the accelerometer can be moved slightly forward. Failing this the degree of stability can be improved if the rate gyro signal is mildly phase advanced, using say a 3 to 1 network.

It is appropriate that the use of compensating networks should now be discussed. Phase advance and phase lag networks are tuned networks and are effective only if the time constants are related to the open loop cross-over frequency. If the open loop gain is likely to vary considerably then a compensating network may improve the response at one gain extreme and actually degrade it at the other. In general it is not feasible to phase advance the error signal as the input to the autopilot has already just been phase advanced in the main guidance loop. The relatively poor signal-to-noise ratio at this input is in no way improved by differencing it with the auto-pilot feedback signals. Superficially it is quite safe to phase advance an accelerometer signal as the error signal has been filtered by the servo and the airframe transfer function. The filtering effect of the airframe is however surprisingly little at the higher frequencies due to the presence of the two zeroes in the transfer function. This is merely saying that even if the airframe does not turn at all as a result of rudder movement there will always be a lateral acceleration due to the normal force from the rudder. Moreover, additional movement of the airframe can result from aerodynamic or propulsion noise. If at any point on the airframe there is a component of lateral movement y_n given by

$$y_n = \hat{Y}_n \sin \omega t$$

then the acceleration at this point is ω^2 times this value. If there are any high frequencies present in this noise then the accelerometer output can be very noisy indeed. Even if the accelerometer can be placed precisely at a node of the principle bending mode it will not be at a node of any of the higher frequency bending modes. So accelerometer signals are rarely phase advanced. Angular rate signals are not likely to be so noisy and angular position signals should be well filtered; signals from position gyros are usually phase advanced.

We conclude therefore that instrument feedback can improve the response of the airframe and that moderate changes in aerodynamic gain can be accommodated. We anticipate however that if really large changes in aerodynamic gain are to be catered for then some form of adaptive system will be required.

6.4 A DISCUSSION OF THE "IMPORTANT" AERODYNAMIC DERIVATIVES

It has been shown that we can readily calculate a required value for y_β and hence y_v if we know the maximum g requirements and the maximum body incidence permissible. Now there are six main lateral aerodynamic derivatives for each plane and we have already dismissed y_r and z_q as being just about negligible. Considering the yaw plane only, we have already discussed and determined y_v. This leaves four others to reconsider. Open loop the weathercock damping is proportional to $(y_v + n_r)$, see Fig 6.3-1. Closed loop the dominant mode damping would be very low indeed without any synthetic damping terms in the feedback path, and yet we have found that it is possible to obtain a well damped closed loop response. The reader who has evaluated the individual terms which make up the denominator coefficients a_o, a_1 etc in the example will realise that the actual value of n_r is not important at all; it can be set to zero and the response, closed loop, is virtually unchanged. This now leaves us with three lateral derivatives n_v, n_ζ and y_ζ. In the previous example we somewhat arbitrarily stipulated a given mean static margin. For a given y_v this fixed the value of n_v (or n_β) and we realised that n_ζ would have to be slightly greater than n_β. Since the rudder moment arm was known we could write down the value of y_ζ once n_ζ was determined. Hence we are really saying that a minimum value for y_v is easily determined and if we set the static margin to a given figure all the necessary design values of the other derivatives automatically follow. Suppose now the static margin is set to half the original assumed value. Does this mean that n_ζ and hence y_ζ can be reduced to half their original values also? The answer is decidely no! The value of n_ζ is a measure of the initial angular acceleration of the body per unit rudder. If this is reduced the autopilot bandwidth will be reduced. So the key lateral aerodynamic derivatives are y_v and n_ζ; all the rest are relatively unimportant. Fig 6.4-1 shows the response of an autopilot with an airframe which, open loop, would be well and truly unstable. The gains etc are the same as those used for the previous worked example with U = 500 but the value of n_v has been changed from 1.0 to -1.0; the accelerometer has been moved to 0.7m forward of the c.g. One sees that, for the same input demand, the rudder finally settles to a steady deflection of opposite sign to that which obtains with a stable missile. The aerodynamic gain increases with an unstable missile, of course, since although the moment from the rudder opposes the unstabilising moment from the body, the forces due to the rudder and the airframe considered as a whole act in the same sense.

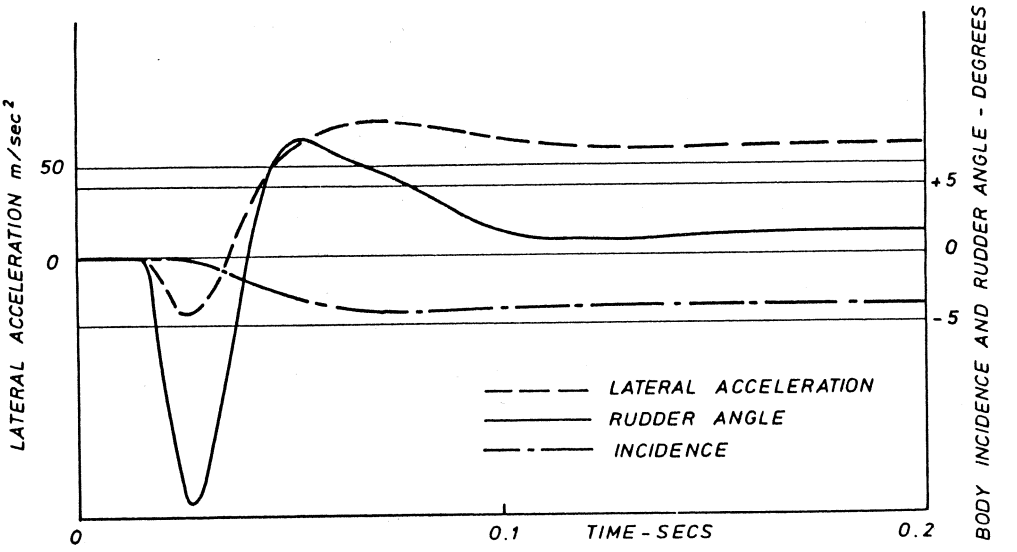


FIG 6.4-1 Response of lateral autopilot which is unstable open loop

The response of a well designed lateral autopilot therefore is relatively insensitive to the position of the c.p. or c.g. (within reasonable limits!) The ideal state of affairs, it is suggested, is for the nominal static margin to be zero. A shift in c.g. will almost certainly occur due to propellant usage; in the majority of missiles the c.g. of the propellant is somewhat to the rear of the missile. The c.p. will change with incidence and with Mach number, the amount depending on the design. It is comforting to know that a satisfactory response is possible with variations of this sort. The condition that one must avoid is too large a static margin. The rudder will have a limited useful travel and a large static margin will make it impossible for the rudder to produce the necessary body incidence to manoeuvre the missile.

At the risk of over-simplifying a rather complicated situation consider once again the denominator coefficients of the autopilot transfer function.

$$a_o \simeq k_a U\,(n_\zeta y_v - n_v y_\zeta) \simeq k_a\, U\, n_\zeta y_v$$

$$a_1 \simeq -k_g n_\zeta$$

$$a_2 \simeq 1/k_s + k_a\,(-y_\zeta - \sigma n_\zeta)$$

$$a_3 \simeq 2\mu_s/k_s\omega_{ns}$$

$$a_4 = 1/k_s\omega_{ns}^{\,2}$$

For low aerodynamic gains the autopilot bandwidth has been found to be relatively low; in such cases the servo dynamics become negligible in which case a_3 and a_4 can be neglected. Also the two terms $-y_\zeta$ and $-cn_\zeta$ tend to cancel out so making another crude assumption that they cancel completely and dividing a_o, a_1 and a_2 by $k_a U n_\zeta y_v$ to make $a_o = 1$ we see the new value of $a_2 = 1/k_s k_a U n_\zeta y_v$. Hence the undamped closed loop natural frequency = $\sqrt{k_s k_a U n_\zeta y_v}$; it is also seen that it is the rate gyro feedback which provides the damping. At the higher loop gains prevailing with higher missile speeds these very simplifying assumptions are too crude to be of quantitative value. Nevertheless it is still correct to say that n_ζ and y_v are always the "important" aerodynamic derivatives.

6.5 THE TWO-ACCELEROMETER LATERAL AUTOPILOT

An accelerometer of gain k_a placed at a distance c ahead of the c.g. with its input axis parallel to the missile oy axis produces a signal equal to

$$k_a (f_y + c\dot{r}),$$

where f_y is the acceleration in the sense oy of the c.g.
An accelerometer similarly oriented but at a distance d behind the c.g. produces a signal equal to

$$k_a (f_y - d\dot{r})$$

Since the additional component from an accelerometer due to its distance forward of the c.g. has a valuable stabilising influence it follows that it would serve no useful purpose at all to place an accelerometer to the rear of the c.g. Nevertheless three very well known British systems use spaced accelerometers to provide instrumental feedback and the ingenious method of mixing the two signals is this: the forward accelerometer gain is increased to $3k_a$ and the rear accelerometer gain is increased to $2k_a$ but this latter signal is fed back positively! The total negative feedback is therefore

$$3k_a (f_y + c\dot{r}) - 2k_a (f_y - d\dot{r})$$
$$= k_a \{f_y + (3c + 2d)\dot{r}\}$$

This is equivalent to having one accelerometer much further forward of the c.g. However the stabilising terms largely affect the a_2 term in the denominator of the closed loop transfer function and the important term $k_g n_\zeta$ which largely controlled the value of a_1 in the one accelerometer and rate gyro system is now absent. Suppose now the signal from the rear accelerometer is lagged by passing it through a filter having the transfer function

$$\frac{1}{1 + T_{ra}s}$$

Ignore the terms due to the angular acceleration for the moment (i.e. regard the accelerometers both situated at the c.g.); the total negative feedback is therefore

$$3k_a - \frac{2k_a}{1 + T_{ra}s}$$

Rearranging this is

$$k_a \frac{(3 + 3T_{ra}s - 2)}{1 + T_{ra}s} = k_a \left(\frac{1 + 3T_{ra}s}{1 + T_{ra}s}\right)$$

and this is equivalent to a 3 to 1 phase advance network. The filtering part of the phase advance has the effect of making the autopilot fifth order which makes detailed analysis of the closed loop rather more tedious. Fig 6.5-1 shows the response of an autopilot to a step demand using two accelerometers

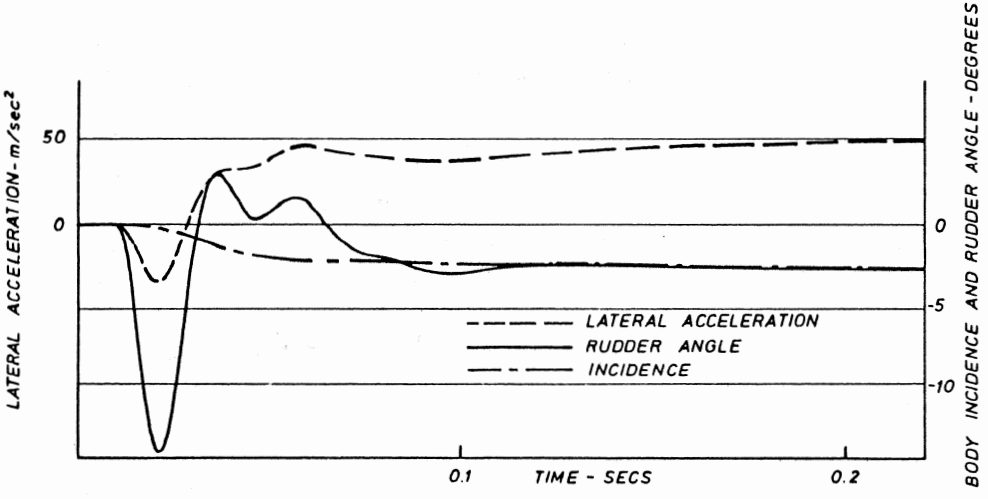


FIG 6.5-1 Response of a two accelerometer lateral autopilot to a step input demand

each spaced 0.5m from the c.g. with the same servo gain and aerodynamic derivatives used for the single accelerometer rate gyro case and U = 500; T_{ra} has been set to 0.038 secs. There are other ways of mixing the signals from the two accelerometers which make this configuration very versatile; in particular it is comparatively easy to produce a quick and well damped dominant mode with zero overshoot. This aspect would be attractive in a system using ram jet propulsion as a large overshoot in incidence has a

tendency to choke the air intake and extinguish the engine.

6.6 THE SINGLE RATE GYRO LATERAL AUTOPILOT

As far as is known only one British system - a medium range homer - uses rate gyro feedback only. Since a rate gyro (used in conjunction with an accelerometer) provides only a small fraction of the main feedback it follows that if one is to design a system with a comparable open loop gain one will have to increase the rate gyro gain considerably, typically about 10 times. This means that the main damping term is now much too large so one can remedy this by passing the rate gyro signal through a phase lag circuit. It is difficult to obtain a reasonably consistent response with moderate variations in aerodynamic gain. Also since the effective d.c. acceleration feedback is proportional to $1/U$ the lateral acceleration/unit input will be proportional to U. It is insensitive to U with accelerometer feedback. Fig 6.6-1 shows the response of a single rate gyro autopilot with lagged feedback and the same loop gain etc as the accelerometer and rate gyro autopilot.

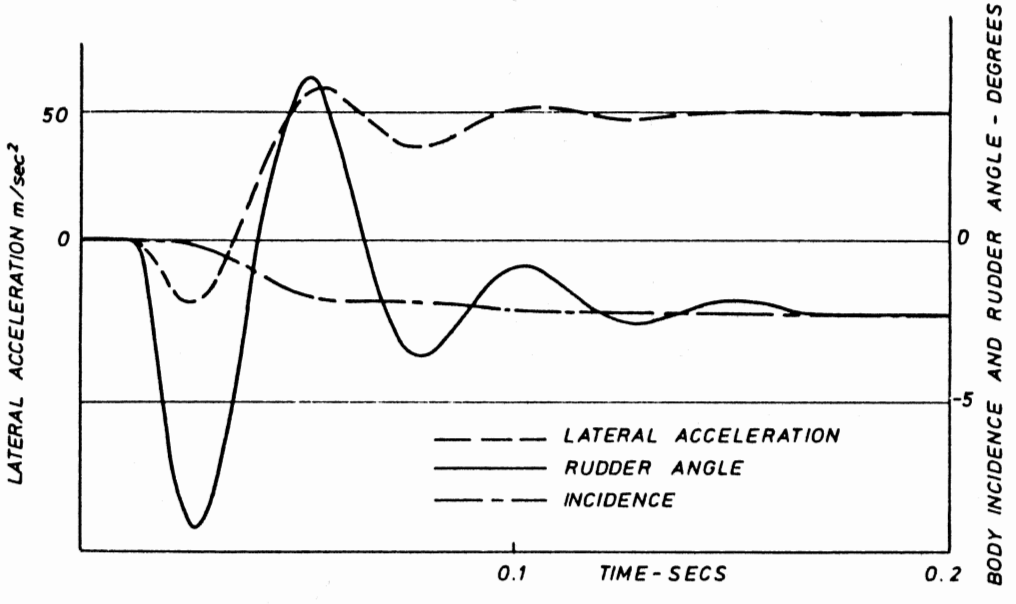


FIG 6.6-1 Response of a single rate gyro lateral autopilot to a step input demand

6.7 THE EFFECT OF CANARD CONTROLS ON LATERAL AUTOPILOT DESIGN

Consider a missile with the same body and main lifting surfaces as our missile with rear controls. We remove these rear controls and install them as canard controls. This moves the c.p. forward so we move the main lifting surfaces back so the static margin is unaltered! The only fundamental change in the aerodynamic derivatives is that n_ζ will now be positive. Consider therefore the design of a lateral autopilot with canard controls with the gains and the magnitude of the aerodynamic derivatives the same as those in the autopilot example is section 6.3. Since a positive rudder movement now produces a steady state positive lateral acceleration the fin servo gain must now be made positive also, but this is a minor detail. If the missile is statically stable the aerodynamic gain will now be larger since the normal force from the rudder and that from the main airframe work together, not in opposition. There is however a noticeable difference in the degree of stability when the loop is closed using accelerometers. If the reader carefully checks through the signs in the new closed loop transfer function he will find that the value of the coefficient a_2 is given by

$$a_2 = 1/k_s + k_a\,(y_\zeta + cn_\zeta), \text{ very nearly.}$$

These three terms are now all positive, and hence the value of this coefficient tends to be significantly larger with a canard control configuration. We have seen (equation 6.3-6) that to maintain stability it is vital to keep the value of a_2 above a certain minimum. This is a direct result of the rudder and airframe working together; the initial feedback resulting from a rudder movement is now negative and not positive as obtains with rear controls. This inherent superior stability with canards when operating closed loop makes it easier to design a satisfactory autopilot in a variable aerodynamic environment. Fig 6.7-1 shows the response of a lateral autopilot with U = 500 m/sec; the only differences between this autopilot and the one whose response is shown in Fig 6.3-3 are that n_ζ has been made positive and the rudder servo gain has been changed in sign.

6.8 A VELOCITY CONTROL AUTOPILOT

The concept of "velocity control" is associated with manual control systems where the operator tracks both the missile and the target and attempts to bring target and missile into alignment by issuing commands to the missile, usually by means of a specially designed thumb controller and a command link to the missile. If no autopilot is used then a constant demand - either up-

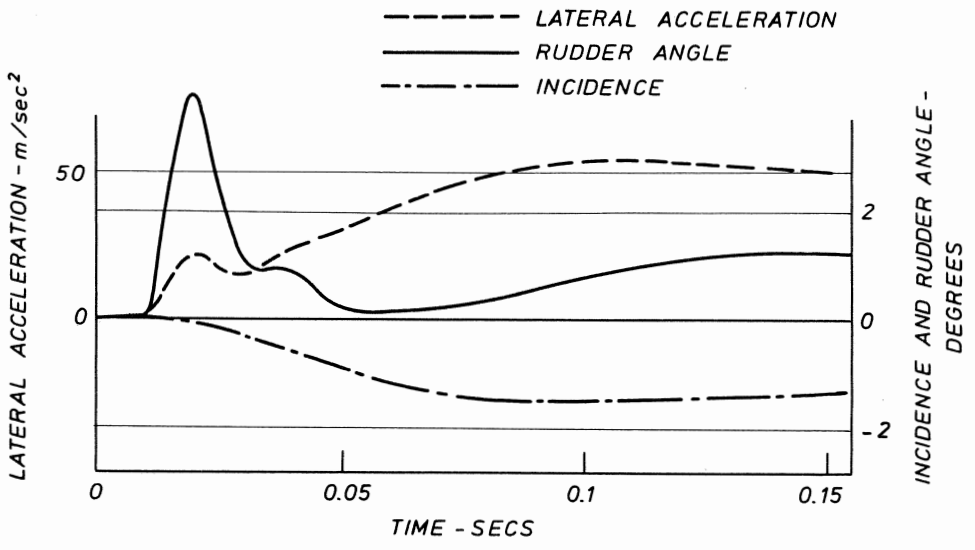


FIG 6.7-1 Response of lateral autopilot with canard controls to a step input demand

down or left-right - will result in the missile developing a steady lateral acceleration. Consider the action necessary by the operator to make a correction in position in the yaw plane. To put in a left correction a good operator will put over his controller to the left until half the "error" is eliminated and will then reverse his controller. Transients in the missile ideally should be so short that the operator is not really aware of their presence. Fig 6.8-1 shows the missile developing a constant lateral

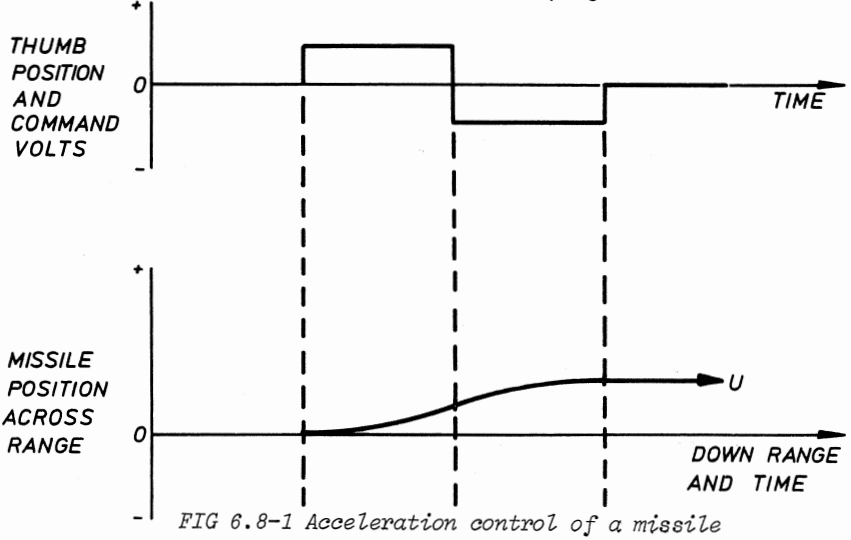


FIG 6.8-1 Acceleration control of a missile

acceleration as a result of operator commands; transients have been ignored. It is clear that a great deal of anticipation and training is required to develop such a skill; the operator is using his own intelligence to supply the necessary amount of phase advance.

Consider a missile carrying a position gyro which feeds back a voltage proportional to yaw angle. The situation now is that the nett signal into the rudder is the difference between the command and the achieved angular position of the missile. When the missile is pointing in the desired direction the rudder will automatically be returned to the central position and the missile will continue to fly straight in this new direction. If this departure from the new flight path direction is a small angle $\Delta\psi_f$ then to a stationary observer the missile appears to have a constant velocity perpendicular to the original LOS equal to $U.\Delta\psi_f$; this lateral velocity is proportional to the command. Such a system of controlling the trajectory of a missile is called "velocity control". Fig 6.8-2 shows a plan of the flight

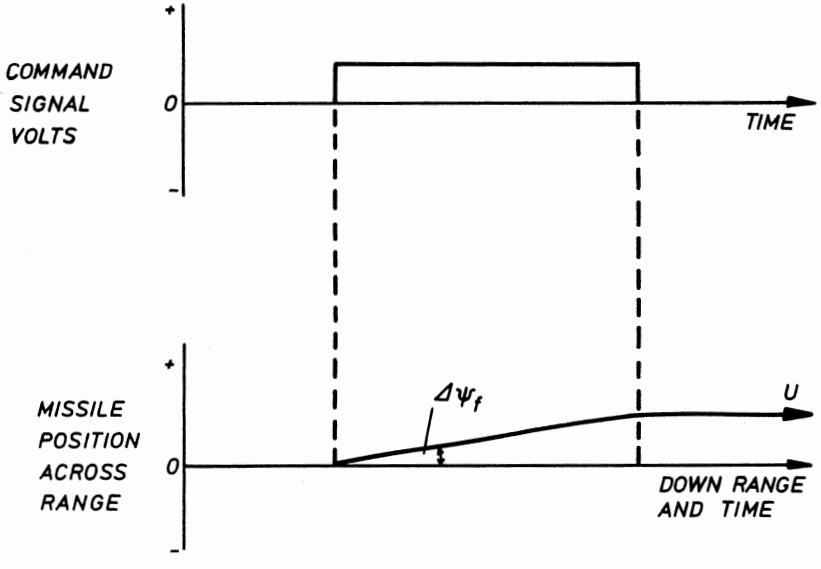


FIG 6.8-2 Velocity control of a missile

path of the missile as a result of a step demand when velocity control is employed. In this country the opinion is widely held that the training of operators is much easier with velocity control than with acceleration control. Nevertheless most people find it comparatively easy to steer a car through gate posts. Does not a car travel in a circle in response to a wheel or "stick" movement and is this therefore not acceleration control also?

Strictly no. The difference is that the guided weapon operator is stationary whereas the car driver has a constant velocity horizon if the steering wheel is over in a fixed position, so the system appears to the driver as a velocity control system. The task of steering a remotely controlled model car through chairlegs is much more difficult; this is a true acceleration control system. Another advantage of a gyro in the loop is that the missile will automatically react to a wind gust, thrust misalignment or disturbance due to a rigging error. With anti-tank missiles there is always the danger of hitting the ground. When a position gyro is used in the pitch plane the missile is usually launched at a set elevation with the gyro uncaged at an angle to the missile framework so that it will fly in a horizontal plane with a small permanent incidence to overcome gravity. Without the gyro gravity compensation still has to be obtained. Any error in biassing the elevator servos results in the missile accelerating upwards or downwards and this calls for more skill from the operator if he is to bring the missile back on course. A velocity control system is suitable for systems designed to hit stationary or slow moving targets where small adjustments only are required by the operator. LOS systems designed to hit fast crossing targets require the missile to execute a curved trajectory and a directional autopilot would merely hinder the operator in trying to keep the missile in the LOS.

Before leaving this subject the effect of the incidence lag should be discussed. A block diagram of a velocity control autopilot is shown in Fig 6.8-3. The transfer function for r/ζ (see equation 4.6-8) can be

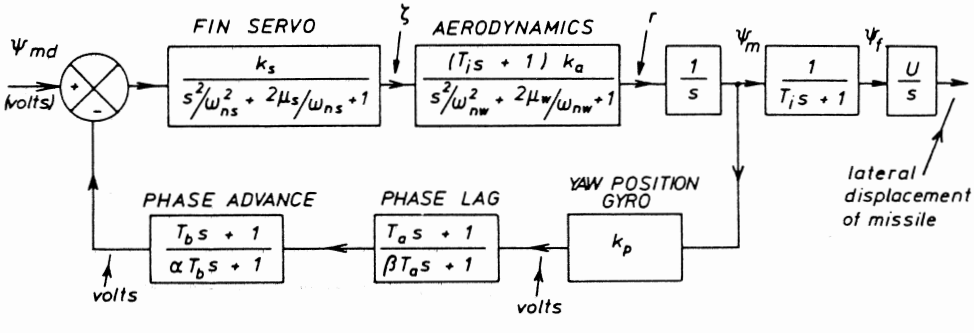


FIG 6.8-3 A velocity control autopilot

rearranged by dividing the numerator and denominator by $(y_v n_r + Un_v)$, the undamped weathercock frequency ω_{nw}. The missile selected has an incidence lag of 0.75 secs and is an anti-tank missile sustaining at a steady speed of

140 m/sec. We have already established that trajectory or flight path rate can be regarded as body rate lagged by this time constant, the incidence lag, see Fig 4.6-2. Now transfer functions assume zero initial conditions so integrating both these quantities gives us the result that flight path direction is related to body direction by the same time constant (the small terms which occur in the full expression for $\dot{\psi}_f/r$ have been omitted for clarity as they hardly affect the argument). Since the operator is trying to change the position of the c.g. of the missile the angular position of the missile is irrelevant; the flight path direction is all important. If therefore we produce a "conventional" type of response for ψ_m/ψ_{md} then the operator will be aware of the considerable lag in producing a change in flight path. For a given incidence lag one method of improving the response is to use a phase lag network in the feedback path. This delays the feedback and allows a large initial overshoot in body angle; if the time constants are chosen carefully stability can actually be improved. The resultant large zero in the closed loop transfer function largely offsets the incidence lag. Fig 6.8-4 shows the response of a velocity control autopilot with ω_{nw} set to 10 rad/sec and the open loop gain to 35. The compensation transfer functions in the feedback path are defined by

$$\frac{0.068s + 1}{0.68s + 1} \times \frac{0.1s + 1}{0.01s + 1}$$

If such an exaggerated body overshoot is not permissible then the incidence lag can be reduced by increasing y_v (this means larger lifting surfaces) or by strongly phase advancing the signal from the operator.

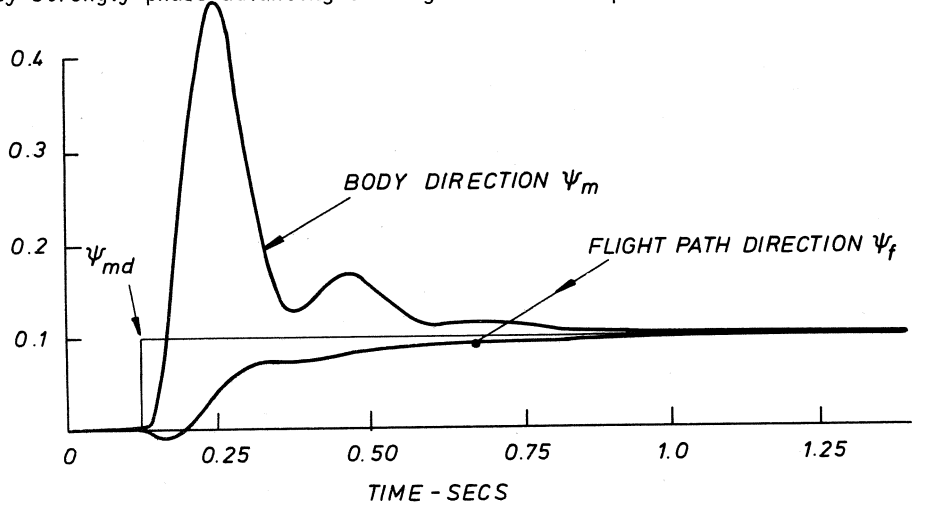


FIG 6.8-4 Response of velocity control autopilot to a step input demand

6.9 LATERAL AUTOPILOTS AND DISPERSION AT LAUNCH

All system biases or disturbances tend to cause inaccuracy but in a closed loop system the ill effects are usually kept fairly small due to the "stiffness" of the closed loop. However some command systems are designed with trackers which may be up to 100m from the launcher. The missile has to fly open loop for some time before it enters the tracker beam. If the dispersion is too great a complete mission failure may result. Even semi-active homers which do not have to enter a beam are required to fly fairly straight during the boost period as during this initial period a rear reference signal may not be available due to excessive flame attenuation from the boost motor. Many modern anti-tank missiles are fired out of a tube which is mounted on a tripod. This very short "gun barrel" will help reduce dispersion. Despite the fact that many command systems use a wide gathering beam which is subsequently narrowed considerably, the question of dispersion at launch is invariably considered and simulated in great detail.

The most serious cause of dispersion is usually thrust misalignment. Even if a convergent-divergent rocket nozzle is perfectly manufactured and aligned, deviations of the gas flow from truly axi-symmetric conditions can occur resulting in an effective misalignment. Darwell and Trubridge (3) have shown how nozzle shapes giving zero gas misalignment can be determined. Nevertheless, this still means that there is a practical limit on the manufacturing tolerances that can be met and checked during inspection. A tolerance of ± 2 milliradians thrust misalignment in any one plane has been quoted for several systems. This may seem a very small misalignment (about 7' of arc) but unfortunately during boost the propulsion forces are large. Suppose we boost at 40g. The lateral component of force will produce a mere $40g/500 \simeq 0.8$ m/sec^2 lateral acceleration for 2 milliradians misalignment. However, thrust misalignment produces a turning moment on the missile also. Consider the situation when the missile body has been turned through a small angle, say 0.05 rad. The longitudinal acceleration of the missile will have a component equal to $40g/20 = 2g \simeq 20$m/sec^2 perpendicular to the original launch direction. At this stage the angular rate will be small so the opposing damping moment will be small. Hence, if no autopilot exists the only way to reduce dispersion due to this cause is to increase the static margin.

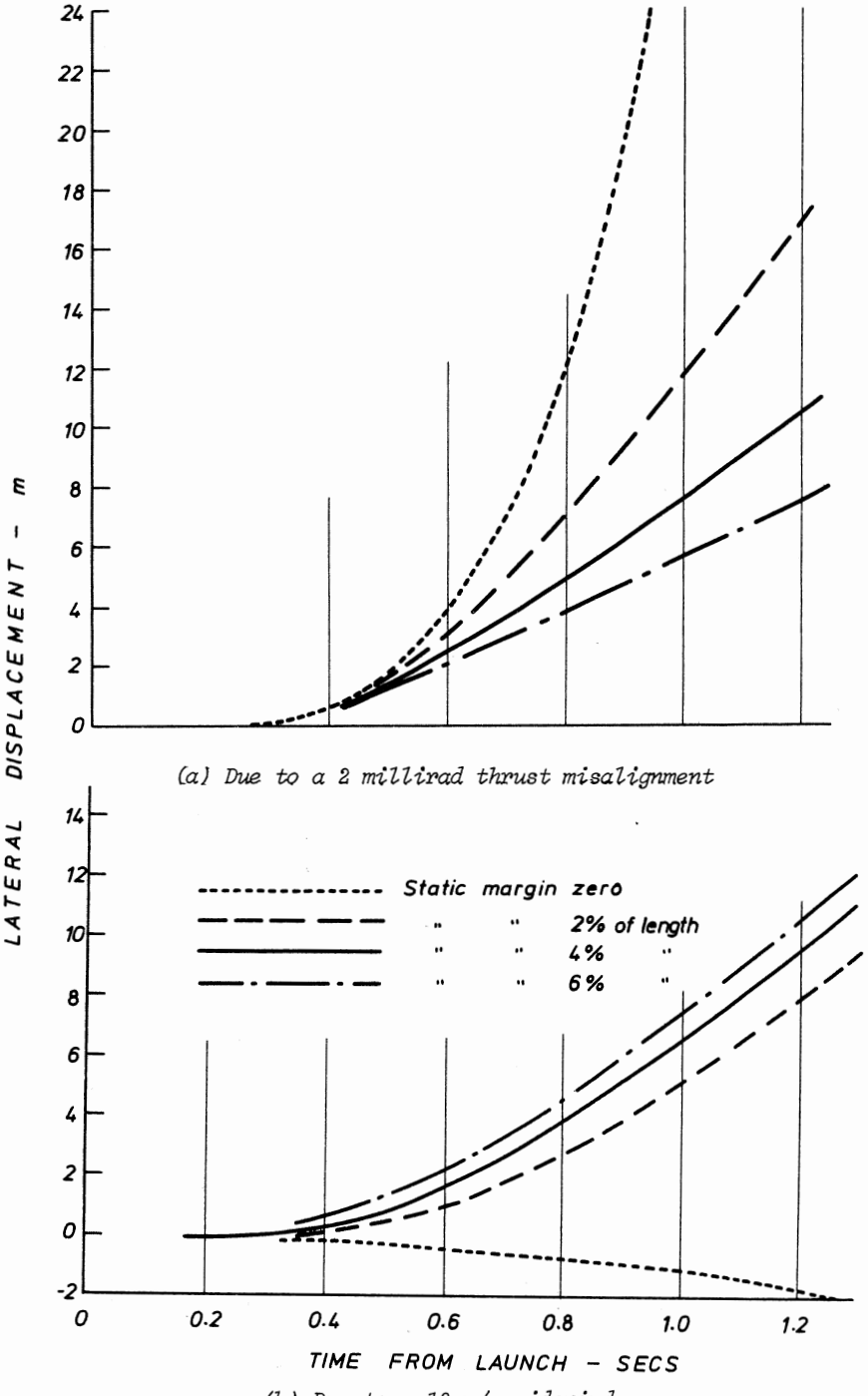
Now consider the effect of a sidewind. If the static margin is zero a sidewind will result in some effective incidence. This will in turn result in a normal force and the missile will eventually drift downwind at wind velocity.

However, a stable missile will turn into the wind and this will have the effect of redirecting the propulsive forces up-wind; this effect is usually far more significant than the natural aerodynamic tendency to drift down-wind. Hence to reduce dispersion due to sidewinds a very small static margin is desirable.

All missiles must exhibit some rigging error and some fin servo bias. The effects of these in causing dispersion could be enormous but with modern production and inspection techniques these effects can usually be kept small when compared with the effects of thrust misalignment and sidewind; they will not be considered in detail here.

Figs 6.9-1(a) and (b) show lateral displacement against time for a sidewind of 10 m/sec and a thrust misalignment of 2 millirad. The missile longitudinal acceleration has been taken as constant at 400 m/sec^2 and the aerodynamic characteristics are the same as the one considered in section 6.3. It has been assumed that the aerodynamic normal force $\alpha \frac{1}{2}\rho U^2$, and the static margin has remained constant during the boost phase. It is seen that provided there is a significant static margin, the dispersion due to a sidewind is not very sensitive to the size of the static margin; the only difference is that the missile initial angular acceleration is greater with a large static margin. The very small dispersion shown as negative with zero static margin is of course a dispersion down wind. When estimating inaccuracies of this nature one often assumes that all biases and disturbances etc have a normal distribution and that none of them are correlated. One then estimates 1σ or 2σ values for each "input" to the system and calculates the likely total effect in the same way as with tracking accuracy in chapter 1. A value of 10 m/sec has been taken as a 2σ value for a cross-wind. Since wind could blow in any direction relative to the launch direction the probable component of a wind of velocity U_w in any one direction (in either sense) is $2U_w/\pi$. Hence we are saying that the 2σ value of the wind in any direction is 15.7 m/sec. Since it is realistic to work on probabilities one's estimate of the wind speed will be influenced by its likely deployment; a system which will be mounted on a ship is expected to experience greater wind speeds than one which is nearly always deployed well inland.

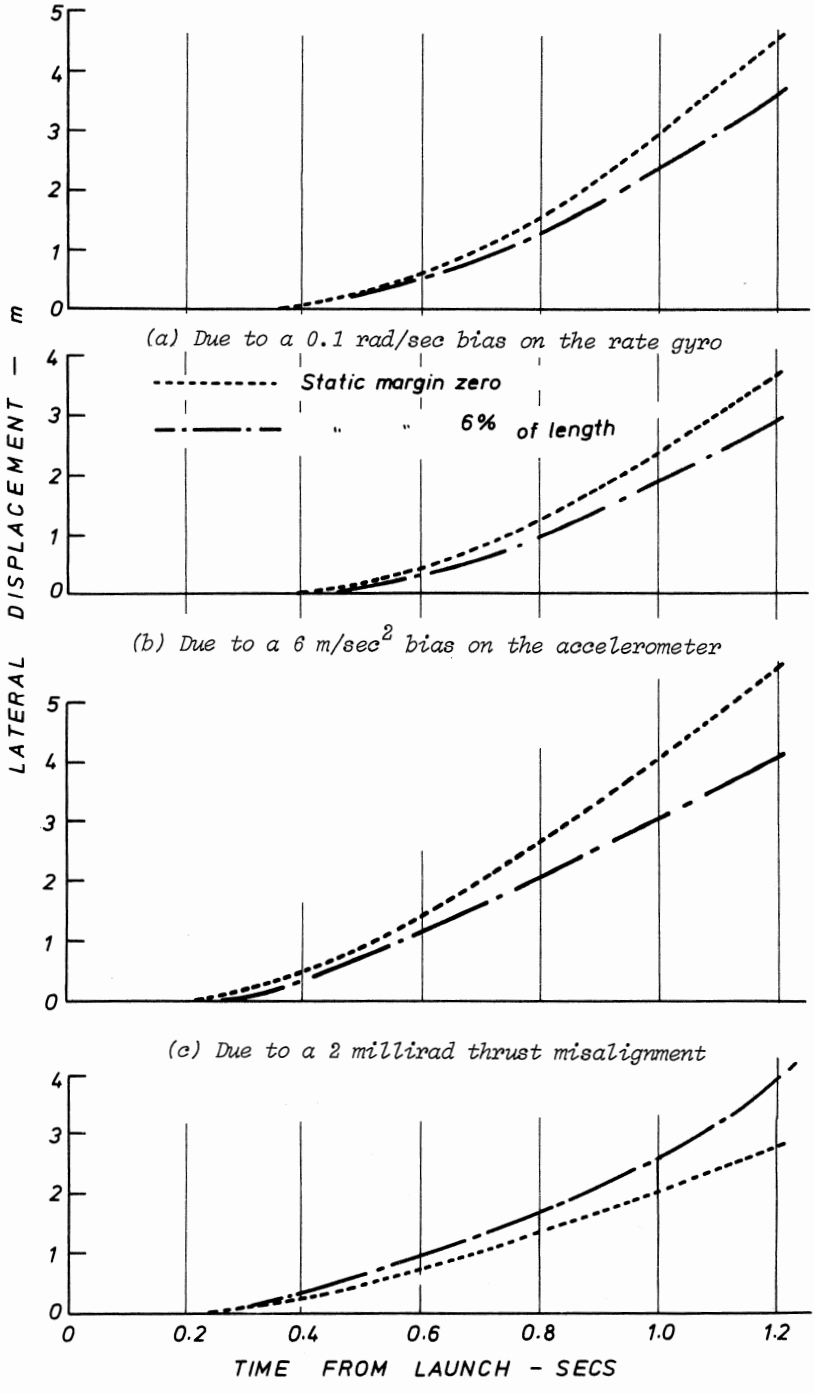
Open loop, dispersion due to thrust misalignment is reduced by a large static margin. It is quite common practice to regard inaccuracies due to tolerances in manufacture to have a normal distribution with zero mean and a 2σ value equal to the tolerance. $\pm$ 2 millirad has been taken as the tolerance.



(a) Due to a 2 millirad thrust misalignment

(b) Due to a 10 m/s sidewind

FIG 6.9-1 Dispersion with no autopilot



(a) Due to a 0.1 rad/sec bias on the rate gyro

(b) Due to a 6 m/sec^2 bias on the accelerometer

(c) Due to a 2 millirad thrust misalignment

(d) Due to a sidewind of 10 m/sec

FIG 6.9-2 Dispersion with an autopilot

Fig 6.9-2(c) and (d) show lateral displacements due to the same causes but now with an autopilot using an accelerometer and a rate gyro as already discussed. The dispersion due to thrust misalignment is seen to be reduced very considerably; one should bear in mind that in practice the static margin during boost is likely to be very small, possibly even negative. The reduction in dispersion due to a sidewind however is disappointing. The behaviour with zero static margin has a logical explanation. If the lateral acceleration due to the sidewind is to the right the autopilot endeavours to steer the missile to the left; it is the large longitudinal acceleration which makes it appear that the autopilot over-compensates.

Unfortunately the inclusion of instruments in the missile control system means that the fin servos are now aware of any instrument biases which would be absent in an open loop system. Figs 6.9-2(a) and (b) show the dispersions due to a bias equivalent to 6 m/sec^2 on the accelerometer and 0.1 rad/sec on the rate gyro. If the full scale deflection of the accelerometer is 30g and that of the rate gyro is 5 rad/sec then these biases (which include any hysteresis effects) represent a tolerance of ± 2% of full scale.

If one accepts these 2σ values there appears to be no great virtue in trying to point the launcher to extreme accuracy. A tolerance here of ± 1^o seems reasonable. The total 2σ dispersion at 1 second after launch when the missile has travelled 200 m downwind for the missile with a 2% static margin is given by

$$2\sigma_t = \sqrt{3.74^2 + 2.56^2 + 2.62^2 + 1.99^2 + 3.49^2} = 6.6\text{m}$$

If the two accelerometer configuration is used then for a comparable performance the biases from these instruments have to be lower. This is because increasing the gain on one accelerometer by 3 and the other by 2 increases the probable total bias by a factor of $\sqrt{3^2 + 2^2} = 3.6$ to 1. Set against this the bias from the rate gyro does not exist. Also this configuration is less effective in producing feedback at low missile velocities, and this is just when it is really wanted. If the missile flight path is changing at a given rate an accelerometer will give a large signal at a high missile speed and a low signal at a low speed ($f_y = \dot{\psi}_f . U$). However, if the body turns, a rate gyro will produce a signal independent of missile speed. Rate gyro feedback is extremely useful in the early boost phase.

Finally it should be noted that if a missile uses thrust vector control the value of Y_ζ and N_ζ and hence y_ζ and n_ζ are not dependant on the speed of the missile. In order to resist a turning moment due to thrust misalignment

say one relies on the product of $N_\zeta\zeta$. With aerodynamic control the value of N_ζ is very low at low speeds and this is the main reason why the dispersion with the autopilot used in the example is surprisingly large. Better results can be obtained with thrust vectored systems. The product $k_a k_s n_\zeta$ is in the feedback path if the "input" is a thrust misalignment or a sidewind, and therefore to a first approximation the lateral acceleration will be inversely proportional to this product. Conversely it is in the feedforward path if the input is an instrument bias and therefore in this context the effect on dispersion is much less.

6.10 AUTOPILOTS FOR ROLL CONTROL

In chapter 3 we discussed some of the reasons why it may be desirable or essential for a missile to be roll position controlled. If four independent fin servos are used then control can be obtained in roll pitch and yaw. Consider an air to air homing missile which is roll position stabilised and due largely to the variability in the launch speed can have a velocity in the range M = 1.4 to M = 2.8. Table 6.10-1 shows the variability of the roll derivatives, aerodynamic gain and time constant over this range for a missile at a height of 1500m. The roll moment of inertia A = 0.96 kg m^2.

TABLE 6.10-1 ROLL AERODYNAMIC DERIVATIVES, GAINS AND TIME CONSTANTS

	M = 1.4	M = 1.6	M = 1.8	M = 2.0	M = 2.4	M = 2.8
$-L_\xi$	7050	8140	9100	10,200	11,700	13,500
$-L_p$	22.3	24.9	27.5	30.3	34.5	37.3
$T_a = \frac{-1}{\ell_p} = \frac{-A}{L_p}$	0.043	0.0385	0.0349	0.0316	0.0278	0.0257
$\frac{\ell_\xi}{\ell_p} = \frac{L_\xi}{L_p}$	316	327	331	336	340	362

The reader is referred to equation 4.6-6 for the definition of aerodynamic gain and time constant. In order to design the roll loop one must know the maximum anticipated induced rolling moment and the desired roll position accuracy. The aerodynamicist estimates that the largest rolling moments will occur at M = 2.8 due to unequal incidence in pitch and yaw and will have a maximum value of 1000 Nm. If the maximum missile roll angle permissible is 1/20 rad then the stiffness of the loop must be not less than 1000 x 20 = 20,000 Nm/rad. This means that in order to balance this disturbing moment we have to use 1000/13,500 rad aileron, and this is approximately 4.2^o. Fig 6.10-1 shows

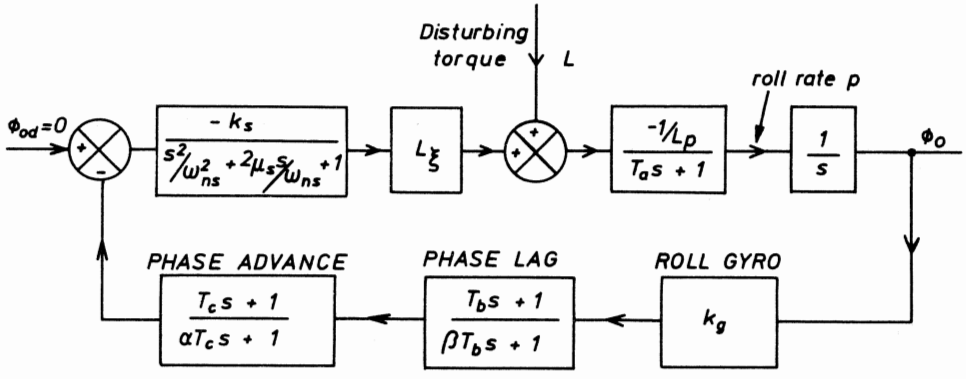


FIG 6.10-1 Roll position control autopilot

the roll position control loop with the demanded roll position equal to zero. Again, the actual servo steady state gain $-k_s$ has to be negative in order to ensure a negative feedback system. Since the steady state roll angle ϕ_{oss} for a constant disturbing torque L is given by

$$\frac{\phi_{oss}}{L} = \frac{0.05}{1000} = \frac{1}{-k_g k_s L_\xi}$$

it follows that $k_s k_g$ must be not less than 20000/13500 = 1.48. If k_g is set at unity then k_s must be 1.48. The open loop gain is now fixed at $1.48\, L_\xi/L_p$ = 535. In order to determine whether the servo dynamics are likely to be significant ignore them for the moment; this means that the loop is now second order with $\omega_n^2 = 535/T_a$ = 20,800 i.e. ω_n = 144 rad/sec. If ω_{ns} = 180 as before the servo will contribute considerable phase lag in the frequencies of interest. With 90^0 phase lag due to the pure integrator and nearly 90^0 phase lag due to the aerodynamic time constant we could use a phase lag compensator to reduce the crossover frequency to reduce the phase lag due to the servo at gain crossover. This can then be followed by a phase advance network. If these transfer functions are set to

$$\frac{1 + 0.05s}{1 + 0.75s} \cdot \frac{1 + 0.0257s}{1 + 0.0018s}$$

then the phase advance numerator cancels the aerodynamic time constant and the Bode diagram for the open loop is as shown in Fig 6.10-2 showing healthy gain and phase margins.

Roll rate control has been used on several missiles. If there are fears that the induced rolling moments could be large and there is no specific requirement for roll position control the maximum roll rate can be very considerably reduced if a rate gyro, orientated so that its sensitive axis is along the

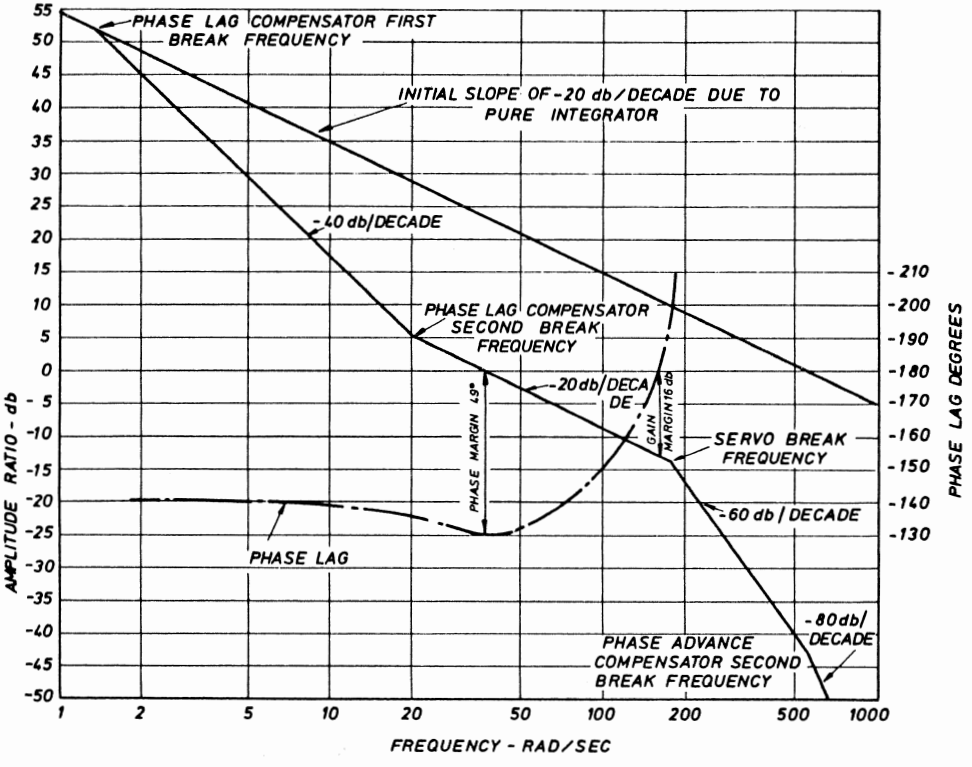


FIG 6.10-2 Bode diagram for roll position control autopilot

missile roll axis, is used to provide feedback into the missile servos such that aileron movement opposes roll rate. In the steady state a constant external rolling moment produces a slow roll rate. This is an attractive idea if the missile carries rate gyros already in the lateral autopilots; position gyros are avoided whenever possible in tactical missiles as they are less robust in design than rate gyros.

A very ingenious method to limit roll rate is to use "rollerons", an entirely passive method. A rolleron consists of a hinged tab mounted on the outer trailing edge of each wing panel. The tab contains a toothed wheel which is driven at a high angular velocity by the impinging air-stream. There is a gyroscopic action if the missile has a roll rate as the tab is caused to precess into the airstream. The resulting torque opposes the missile roll rate. If the hinge line is canted at 45° to the missile transverse axis some damping in pitch and yaw is also achieved.

A roll position demand autopilot is required if twist and steer is used. The method is similar to that used for roll position stabilisation except that a

demand for roll angle from the guidance system now exists. Consider how the two demands R and ϕ will be implemented. In cartesian control elevation and azimuth signals can have algebraic signs i.e. for up - down and for left - right. With polar coordinates neither R nor ϕ have signs. This means that R must be interpreted as an "out" demand at all times. Now consider what would happen if we use a circular potentiometer on the roll position gyro for measuring roll angle. As a potentiometer must have a beginning and an end, some trouble is taken in designing a potentiometer with as large an electrical angle as possible, say 358^{o}. For convenience call this 360^{o} and assume that the total voltage across the potentiometer is 18V. Suppose the missile is at a roll angle of 10^{o} i.e. the demand is 0.5V. Suppose the guidance now sees the target at 350^{o} relative to the missile. The demand must therefore now be equal to 17.5V. In the absence of sophistication the missile will roll through 340^{o} and if it should overshoot beyond 358^{o} the error will appear enormous and it will roll round again, and so on. The system appears impractical.

One way of overcoming this problem is to incorporate logic circuitry. The roll demand is compared with the measured roll position in the usual way but before passing this signal to the aileron servo amplifiers as a true error signal the output is fed to two comparator amplifiers (logic 1 and logic 2) each of which drives a transistor switch. If the error voltage exceeds that corresponding to a change in position of more than $+180^{o}$ logic 1 will operate and add a discrete voltage equivalent to -360^{o} so that the effective demand is less than -180^{o}. Similarly if the "error" exceeds -180^{o} logic 2 will operate and add a discrete voltage equivalent to $+360^{o}$ so that the effective demand is less than $+180^{o}$. For error voltages equivalent to less than 180^{o} neither logic circuit operates. Such a system:

(a) ensures that the missile selects the shortest way round

(b) cannot at any time try to rotate more than 180^{o}.

The discontinuity at the ends of the potentiometer now no longer presents a problem. It is true that the system could oscillate backwards and forwards in this dead zone and therefore in one very small sector there is a possible servo error of 2^{o}. Fig 6.10-3 illustrates this principle. In a sense a twist and steer system seems natural for a homing system because as the missile rolls as a result of a guidance command there is inherent roll angle feedback (the guidance receiver rolls with the missile). Hence there is always inherent unity feedback. In order to achieve closed loop stability therefore the guidance ϕ signal will have to be compensated before passing to the fin servos.

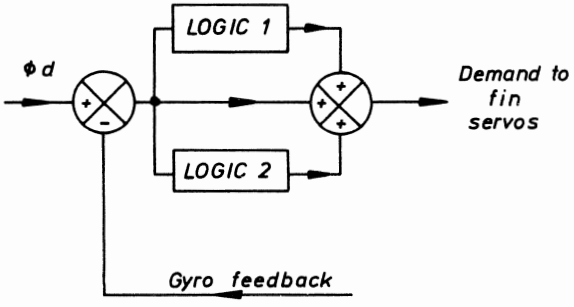


FIG 6.10-3 Input logic for roll position demand autopilot

6.11 THE EFFECT OF ROLL RATE ON LATERAL AUTOPILOT PERFORMANCE

Consider a cartesian controlled missile with two identical lateral autopilots; it is free to roll and the two guidance commands are resolved in the missile as discussed in section 5.6. The resolved commands are usually called "demands for lateral acceleration in freely rolling missile axes". The achieved total lateral acceleration of the missile in space can be regarded as resulting from a lateral acceleration along the missile z axis plus another along the missile y axis. These two vectors will have vertical and horizontal components such that the achieved lateral acceleration in space should ideally be identical to the original demanded lateral acceleration, see Fig 6.11-1.

Now imagine the missile to be rolling at a steady rate p rad/sec, and consider for the moment a constant up demand; the left-right demand is zero. The elevator and rudder servos will be executing simple harmonic motion at a frequency p rad/sec and the two servos will exhibit a phase lag ϕ_s at this frequency. If at a given instant the missile is at zero roll angle then the elevator angle will be reduced by a factor equal to cos ϕ_s and the rudder servos will have a component proportional to sin ϕ_s. Put in another way, which is easier to see, if one waits a short time so that the missile rolls through an angle ϕ_s then the elevator output will now be equal to the demand and the rudders will be centralised. This means that the elevator effort in the missile *oxz* plane is reduced by cos ϕ_s and part of its effort appears in the missile *oxy* plane. The same is true of the rudder servo if there is a demand in the horizontal plane. Hence there is cross-coupling between these planes due to the effect of roll rate and the servo response. This will have an adverse effect on the accuracy of the system as a whole, and this will be

considered in the next chapter but for the moment let us see whether the stability of the two-plane system is affected. Fig 6.11-2 is a block diagram

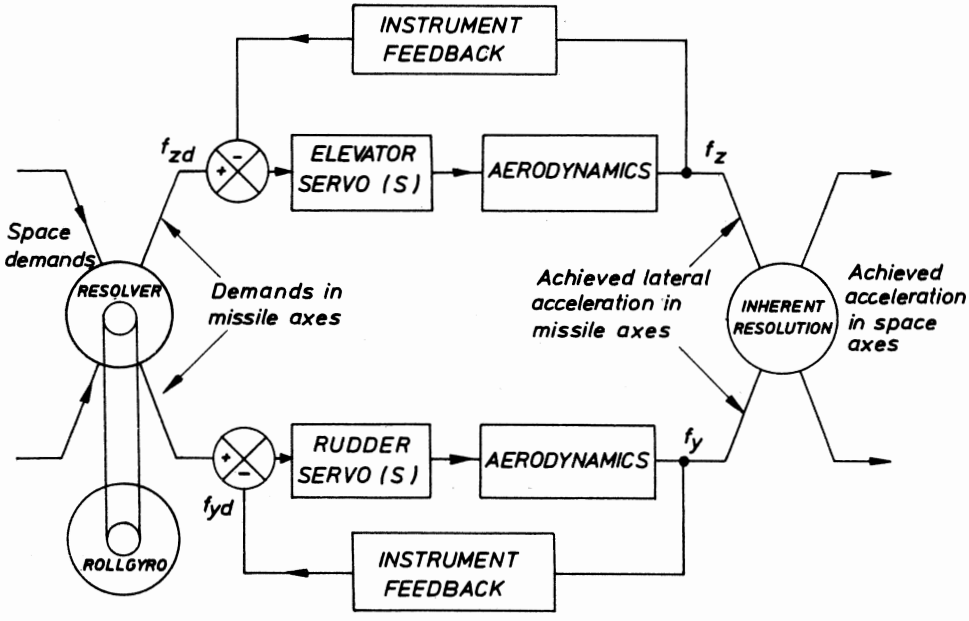


FIG 6.11-1 Resolution of guidance signals

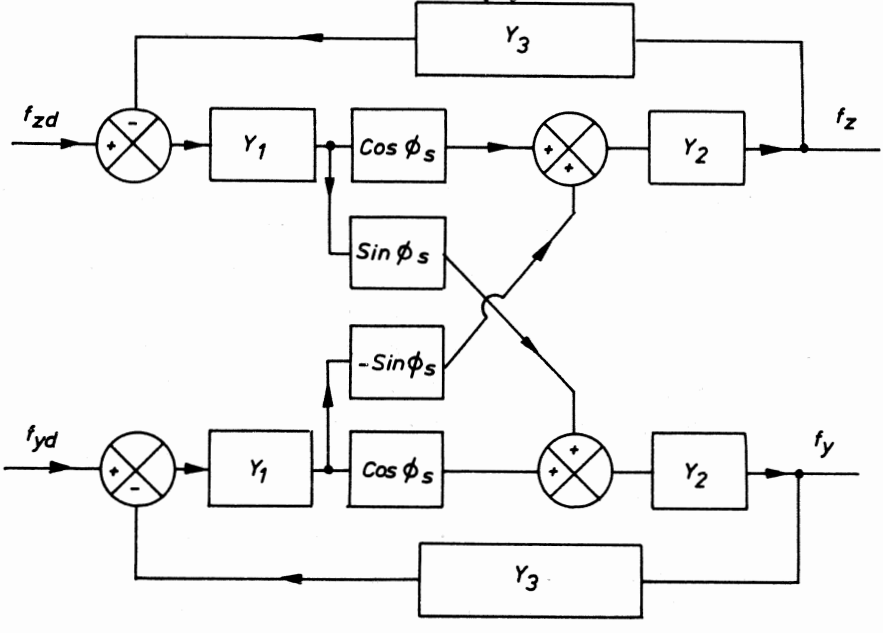


FIG 6.11-2 Cross coupling between two identical lateral autopilots due to roll rate and fin servo lag

of the coupled system with Y_1, Y_2 and Y_3 the transfer functions of the servo, aerodynamics and total effective feedback respectively. We will obtain the transfer function for f_y/f_{zd} setting $f_{yd} = 0$. The transfer function f_z/f_{yd} will be identical provided that for sin ϕ_s we read - sin ϕ_s. If cross-coupling is ignored then

$$f_y = (f_{yd} - f_y\, Y_3)\, Y_1 \cos \phi_s\, Y_2 \qquad (6.11\text{-}1)$$

$$\text{and } f_z = (f_{zd} - f_z\, Y_3)\, Y_1 \cos \phi_s\, Y_2 \qquad (6.11\text{-}2)$$

The additional signals entering these two loops due to cross-coupling are

$$(f_{zd} - f_z\, Y_3)\, Y_1 \sin \phi_s \text{ and } (f_{yd} - f_y\, Y_3)\, Y_1\, (-\sin \phi_s)$$

If these two signals are multiplied by Y_2 and added to the right hand side of equations 6.11-1 and 6.11-2 respectively then this represents the coupled situation. If $f_{yd} = 0$ and we simplify by writing the uncoupled open loop transfer function $Y = Y_1\, Y_2\, Y_3$ manipulation yields

$$\frac{f_y}{f_{zd}} = \frac{Y_1\, Y_2 \sin \phi_s}{1 + 2Y \cos \phi_s + Y^2} \qquad (6.11\text{-}3)$$

showing that there is no cross coupling if $\phi_s = 0$ and the denominator reduces to $(1 + Y)^2$. The characteristic equation for each plane is of course $1 + Y = 0$. For marginal stability, using Nyquist's criterion

$$2Y \cos \phi_s + Y^2 = -1 \qquad (6.11\text{-}4)$$

For $\phi_s > 0$ the complex roots are

$$Y = -1 \pm \sin \phi_s\, j$$

If ϕ_s is less than about 30^0 we can write

$$Y \simeq -1 \pm \phi_s\, j$$

Considering the root $Y = -1 + \phi_s\, j$ this is saying that for marginal stability the open loop phase instead of being - 180^0 is now - $180^0 + \phi_s$; in other words for small roll rates the autopilot phase margin is reduced by the servo phase lag at the roll frequency. Or more correctly one should remember that in originally assessing phase margin for the zero roll rate case the servo lag at gain crossover frequency was already included in the open loop transfer function Y. Hence one must shift the frequency axis for the servo by an amount equal to the roll rate. If therefore the roll rate is taken at p rad/sec and one is considering a frequency ω in the stability analysis then the servo frequency must be taken as $(\omega + p)$ rad/sec. A fuller treatment is given by Frary (5) who confirms these general conclusions, but by an entirely different approach. This destabilising effect of roll rate explains why most lateral autopilots are designed with a fairly large phase margin at zero roll rate,

typically 50^{o} or more.

Frary also considers the inertial cross-coupling effects together with the contribution from the Magnus derivatives. The Magnus effect is easily appreciated if one considers a cylinder spinning about its longitudinal axis in still air. If the same cylinder is also moving perpendicular to this axis the surface velocities relative to the airstream on either side of the cylinder are unequal. This produces a difference in the boundary layer thickness resulting in a force at right angles to the direction of motion. Hence, if a missile is rolling and has a component of lateral velocity a force at right angles to this velocity will be experienced. The derivatives are of the form y_{pw} and z_{pv} and the forces are functions of roll rate multiplied by lateral velocity. It is found that for a typical anti-aircraft missile the Magnus terms will appreciably alter the airframe response only for roll rates above 200 rad/sec. Several modern supersonic missiles which are free to roll have been observed to roll at very low rates throughout an engagement; 20 rad/sec would be regarded as an abnormally high roll rate. Hence Frary concludes that the Magnus effect can be safely neglected in the great majority of systems.

The effect of the inertia cross-coupling terms however can be reckonable. In arriving at equations (4.2-12) and (4.2-13) the roll rate was assumed to be small. If the roll rate is large then for a symmetrical missile these equations become

$$\dot{q} - \left(\frac{C - A}{B}\right)pr = \frac{M}{B} \qquad (6.11\text{-}5)$$

$$\text{and } \dot{r} - \left(\frac{A - B}{C}\right)pq = \frac{N}{C} \qquad (6.11\text{-}6)$$

These new equations can now be used together with the angle equations and force equations to obtain the aerodynamic transfer functions which will be a function of roll rate. Since the lateral moments of inertia B and C are much greater than the roll moment of inertia A the functions in brackets approximate to unity for most missiles. Frary found that the typical surface-to-air missile went unstable open loop at a roll rate of 135 rad/sec; the effect closed loop was however, much less. This will surprise most readers who know that a conventional shell is spun at a very high rate in order to make it aerodynamically stable. However, a paper by Murphy (8) concludes that in certain regimes and configurations a statically stable missile may be made dynamically unstable above a certain spin rate.

The general conclusions are therefore that provided roll rates are not really high then Magnus and inertial cross-coupling effects should be very small.

However, even small roll rates tend to unstabilise a two plane autopilot unless the servo response is really fast.

6.12 AUTOPILOTS AND A CHANGING ENVIRONMENT

We have already discussed the factors that are responsible for the variations in the gains, time constants etc in an autopilot i.e. mass changes due to propellant usage, changes in speed and height and changes in static margin. Mass and inertia changes are relatively small and considerable changes in static margin can be accommodated in a well designed autopilot. However, large changes in speed and height will result in such substantial changes in the important aerodynamic derivatives that it will become essential to vary at least one of the gains in the loop in order to maintain a satisfactory response. Clearly the feedback gain must be kept constant to maintain the closed loop d.c. gain. In practice the feedforward gain is easily controlled by varying the gain of the amplifier which sums the input demand and feedback signals. But to vary this parameter only will rarely be sufficient to maintain adequate damping as well. In the lateral autopilot using an accelerometer and a rate gyro one could alter the effective rate gyro gain; in the two accelerometer case good control over damping can be obtained by varying the rear accelerometer lag time constant. Roll position control autopilots are altogether simpler as the aerodynamic gain L_ξ/L_p will not change at all with changes in air density and is not very sensitive to changes in speed. The aerodynamic time constant however can undergo large changes. A simple gain change may be all that is necessary. If it is a roll position demand loop the closed loop gain will not change at all with aerodynamic variations since it is a type 1 system.

Changes in gains and time constants are easily effected by switching in different values of resistance and capacitance used in conjunction with operational amplifiers. In the case of a surface-to-air system there will probably be a computer associated with the surveillance and guidance system. The velocity of the missile as a function of time will be known within reasonable limits, and the tracker elevation angle together with a knowledge of time and hence missile range will enable an estimate of missile height to be made. Commands from the guidance computer can be made to operate solid state switches in the missile to effect the necessary changes. The number of switching zones is kept to a minimum. Conversely, the missile can be independent of the ground in this respect and can effect the necessary changes

if some measurements concerning the aerodynamic environment are made. Since all normal force derivatives are proportional to $\frac{1}{2}\rho U^2$ and are not very sensitive to Mach number an approximation to the dynamic pressure can be obtained by means of a bellows connected to a pitot tube on one side and the static head on the other. The differential pressure and hence resultant force on the diaphragm is nearly proportional to the dynamic pressure if the range of Mach number is not too great. Since the actual differential force will be small an instrument servo can be used to produce an angular movement proportional to this pressure, see Fig 6.12-1. This angular motion can be

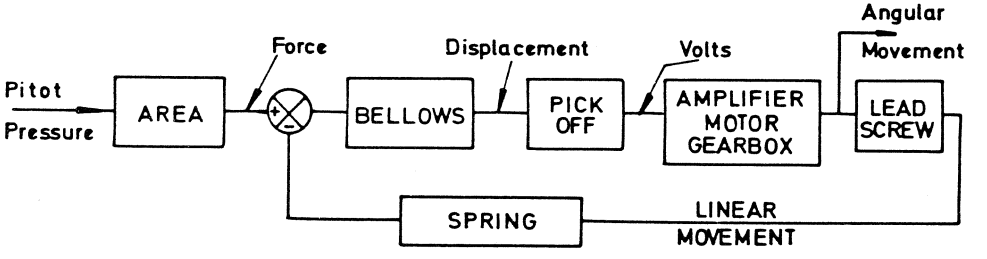


FIG 6.12-1 Instrument servo as part of an aerodynamic gain unit

coupled to potentiometers in which case the changes can be continuous. A more likely method in the future is the use of a dynamic pressure sensor in conjunction with solid state multipliers and switches.

Such autopilots are not truly "adaptive" in the generally accepted sense. Model reference systems are used on many large aircraft and despite their relative complexity the cost of such systems is still small compared with the cost of the aircraft. The cost of such a system has, as far as is known, precluded their use in guided weapons; the simpler system described above, after all, does the job reasonably well.

6.13 AZIMUTH CONTROL BY GYROS AND ACCELEROMETERS

There are many missile systems which use a surveillance system to determine the height and direction of a target. The missile uses homing guidance when it is near to the target, say at a range of 2-5 km and relies on azimuth guidance from a gyro during the mid-course sustain phase. The azimuth autopilot is similar to that used in a velocity control system except that there may be no commands at all. Such systems typically have flight times up to 100 seconds and sustain at a velocity about 250 m/sec; this low velocity results in a greatly reduced drag. One uses a much higher quality azimuth gyro than would be used in an anti-tank weapon with a random drift rate

typically of 5^{o}/hour and a anisoelastic drift rate of 1.5^{o}/hour/g^{2}. Gyros used to keep a missile on course for a comparatively long time are usually called "guidance" gyros. Since the missile homing head has a limited angle of look one has to estimate the maximum likely look-out angle when the missile enters the terminal homing phase. There are many causes of error in such a trajectory:

(a) We may not have aimed correctly in the first place or the target may have moved an unknown amount during the missile flight.

(b) The gyro will drift during boost.

(c) The gyro will drift (much slower) during the sustain phase.

(d) The missile will drift downwind due to a side wind. The azimuth gyro will do nothing to correct this as all it knows is the direction of the body, not the direction of the velocity vector.

If the missile is sustaining at a constant velocity and we are considering long times of flight it is reasonable to assume that a given change in body direction will result in the same change in flight path direction. Thus, if the gyro drifts by an amount ψ the missile flight path will also change by ψ. Let

R = the range from the launcher at the commencement of the engagement to the target at impact.

r = the range from the missile to the target at the start of the homing phase.

ψ_1 = the error in aiming.

ψ_2 = the gyro drift during boost.

ω = the gyro drift rate during the sustain phase assumed constant.

U_w = the component of wind perpendicular to the flight path.

T = the time to reach a range of $R - r = (R - r)/U_m$, see Fig 6.13-1.

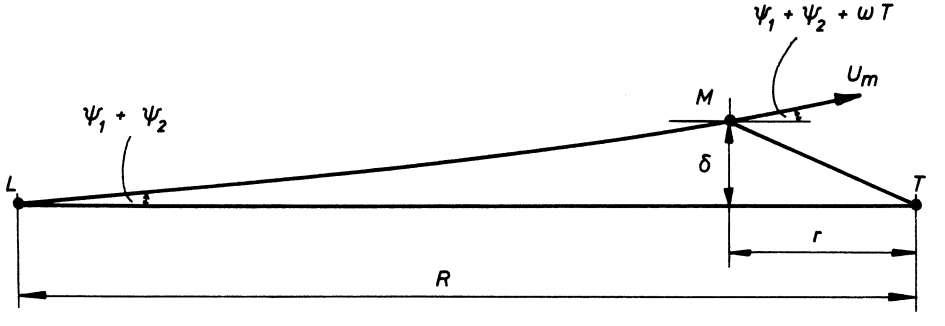


FIG 6.13-1 Geometry of engagement, plan view

Assuming the missile range after boost is negligible and making small angle approximations then the total lateral displacement δ at a range $R - r$ is given by

$$\delta = (\psi_1 + \psi_2)(R - r) + \int_o^T U_m \omega t \, dt + U_w T$$

and the angle of look θ is given by $\delta/r + \psi_1 + \psi_2 + \omega T$. This reduces to

$$\theta = \psi_1\left(1 + \frac{R-r}{r}\right) + \psi_2\left(1 + \frac{R-r}{r}\right) + \frac{\omega(R-r)}{U_m}\left(1 + \frac{R-r}{2r}\right) + \frac{R-r}{r} \cdot \frac{U_w}{U_m} \quad (6.13\text{-}1)$$

Since these four components can be assumed to be uncorrelated one can estimate the probable dispersion if the 1σ values of each are known. It is seen that even if errors due to aiming and gyro drift can be made negligible nevertheless the dispersion due to a sidewind can be considerable e.g. 1 km in a range of 25 km for a sidewind of 10 m/sec and a missile speed of 250 m/sec. The alternative is to use a "cross-track navigator" which employs negative feedback from a very low bias accelerometer mounted on a gyro gimbal such that its sensitive axis is truly horizontal and perpendicular to the missile original heading. It is absolutely essential that this accelerometer does not sense a component of gravity due to being off level otherwise spurious signals integrated over a long flight time could result in a large dispersion. However it may be possible to employ a gyro which is slightly pendulous to prevent drift. In the absence of biases this system has the potential of eliminating dispersion since any sideways component of velocity due to any cause will show initially as an output from the accelerometer. Reference (4) gives some particulars of systems with a cruise phase under instrument control.

6.14 HEIGHT CONTROL AND SEA SKIMMING SYSTEMS

Many countries have designed guided weapon systems to fly low or very low under altimeter control in order to remain below the enemy's radar horizon during most of the flight. This includes remotely piloted vehicles used for surveillance and target designation and anti-ship missiles whose flight path lies wholly over the sea. If one wishes to fly at a height of 100 m or more the accuracy of a barometric altimeter or even a piezo electric pressure transducer is usually good enough to prevent one from hitting the ground or sea. Moreover most land surveillance vehicles are fitted with television cameras which continusously transmit a picture of the terrain back to the operator who has direct control over its height and course; controlling such a vehicle so that it does not hit the ground is not generally regarded as being difficult. Now, even a rough sea is more or less level and in general

it is possible to fly much lower over the sea with the same chance of "ditching". If one is thinking of flying at a height of 10 m or less above mean sea level one must know something about the sea state and one must use a radio altimeter with a very small bias. This section is therefore primarily concerned with the problem of flying very low over the sea solely under altimeter control.

We now consider the sea spectrum. There is a great deal of literature both theoretical and based on actual measurements concerning ocean waves. An early but fundamental text is given by Darbyshire (6). There is a wealth of observed data in "Ocean Wave Statistics" (7) and "Standard Wave Spectra" (8). In the same way that wind speeds are denoted by numbers on the Beaufort scale, so the international sea state is denoted by numbers from 1 to 6. This number is defined by a parameter known as the significant wave height H_{sig} which is the average of the highest third of all the waves recorded in a specified interval. H_{sig} is more usually written as $H_{3}rd$, and the values are invariably quoted as peak to trough values. There is some agreement that a typical sea state can be considered as a random collection of sine waves with a Gaussian distribution of amplitudes and a narrow frequency distribution dependant on the sea state. The relationship between $H_{3}rd$ and the standard deviation σ is given by $H_{3}rd = 4\sigma$ where σ is measured about local mean sea level.

Now the sea state is undoubtedly due to the wind speed but it is not possible to give an exact relationship between $H_{3}rd$ and wind speed as it depends also on how long the wind has been blowing, the distance over which it has been acting (called the fetch), the depth of the water, and whether the sea is affected by swell (due to a wind which is, or has been blowing elsewhere). So although a generally accepted equation concerning sea state will be given, it must be pointed out that it is empirical. The Pierson-Moskowitz spectrum is a standard accepted internationally as a one dimension spectrum for a stationary observer for the case of a fully developed sea in deep water and is given by

$$\Phi_s(\omega) = \frac{0.78}{\omega^5} e^{\frac{-4 \times 0.78}{(H_3rd)^2 \omega^4}} \quad m^2 \text{ rad}^{-1} \text{ sec} \qquad (6.14\text{-}1)$$

It will be noted that

$$\int_0^\infty \Phi_s \, d\omega = (H_3rd/4)^2 = \sigma^2$$

Figs 6.14-1 (a), (b) and (c) have been drawn for different sea states. It is

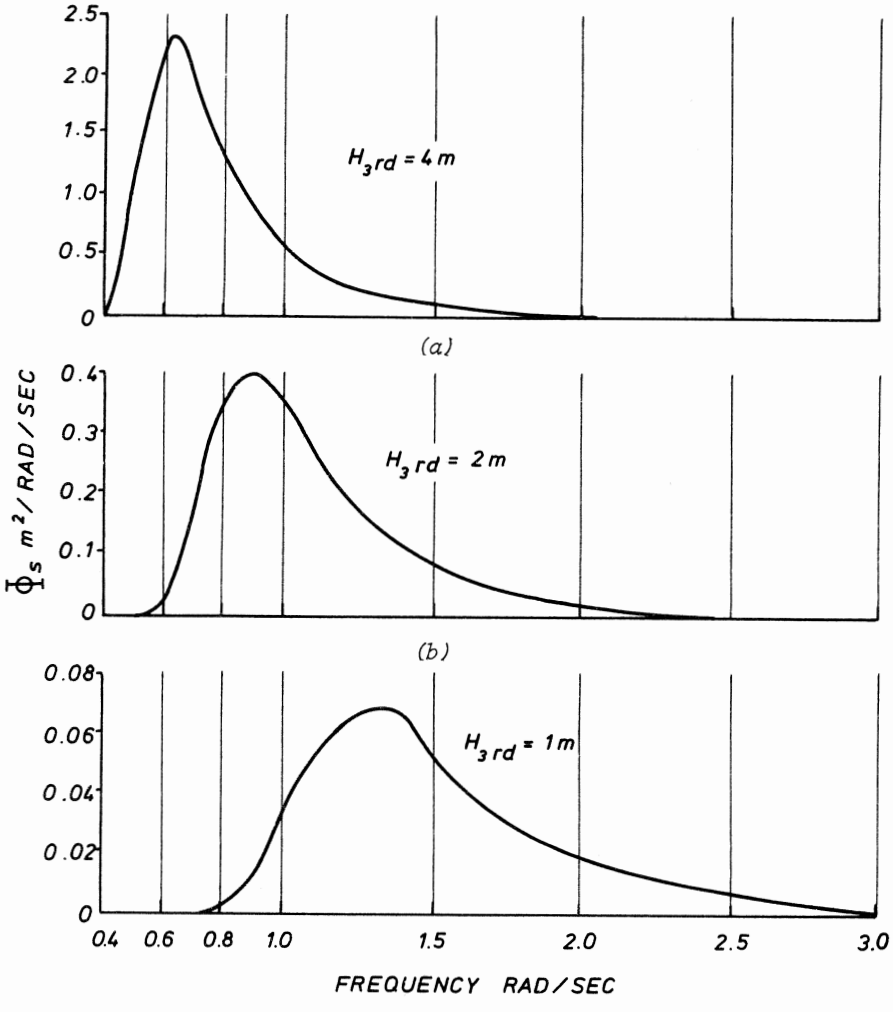


FIG 6.14-1 Sea spectrum for different sea states

interesting to note that according to Hogben and Lamb (7) the most probable value for $H_{3}rd$ off Cornwall for the whole year is 0.75 m and is 1.8 m for the North Atlantic. It is seen from these figures that for a rough sea ($H_{3}rd$ = 4m) the spectrum is narrowly centred about 0.6 rad/sec or thereabouts and this corresponds to one wave every 10 secs. The main frequencies are about double this for $H_{3}rd$ but the spectrum is somewhat wider. If the wave frequency is f_w (Hz), c the velocity of wave propagation and λ_w the wavelength then

$$\lambda_w = \frac{g}{2\pi f_w^2} \text{ and } c = \frac{g}{2\pi f_w}$$

These frequencies are of course for a stationary observer. If the missile velocity is at an angle α to the wave direction the wave frequency as seen by the missile f_m is given by

$$f_m = \frac{U_m \cos \alpha}{\lambda_m} \pm f_w = \frac{U_m \, 2\pi \, f_w^2 \cos \alpha}{g} \pm f_w \qquad (6.14\text{-}2)$$

If, for example f_w is taken as 0.1 Hz and U_m = 300 m/sec, the frequency as seen by the missile could be anything from zero to 2 Hz, but would be up to 3-4 Hz for the shorter wavelengths more characteristic of a calmer sea. If the height lock system approximates to a second order system with an undamped natural frequency of 2π x 3 rad/sec and a damping ratio of 0.5 then the steady state error for a sinusoidal input at this frequency is also a sinusoid whose amplitude is $\sqrt{2}$ times the input amplitude. This is equivalent to flying precisely level with a virtual input $\sqrt{2}$ times the real input amplitude! Since it will be shown that the system bandwidth is unlikely to exceed 4 rad/sec we must regretfully come to the conclusion that it will not be possible to follow the sea profile at all. In other words more often than not the output of the height lock system will be 90^0 or more out of phase with the input and the mean following error is likely to be larger than the mean sea height.
Fig 6.14-2 shows one type of height control system. Accelerometer and rate gyro feedback in the vertical plane ensure a reasonably well damped and consistent

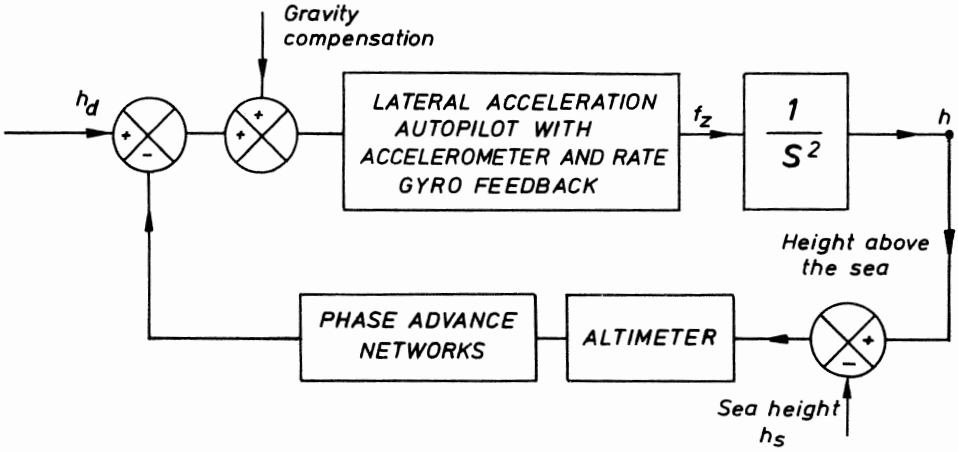


FIG 6.14-2 A height control system

response from the missile. Since height is the double integral of vertical acceleration there are two inherent integrations in the complete system if altimeter feedback is also used. A good deal of phase advance is essential to

ensure loop stability. The problem of loop stability is eased somewhat if the phase lag due to the autopilot at gain cross-over frequency is almost negligible, say 5^{o}-10^{o}. This can be achieved if the loop gain is very low, say 4 or less and the autopilot bandwidth is at least 10 times the loop gain cross-over frequency; this latter frequency would be 2 rad/sec if the loop gain is 4. It may be impractical to obtain a very fast autopilot if the missile is subsonic and is on the big side, say over 150 kg. Since the altimeter sees the difference between the missile height and the instantaneous sea height (which is noisy) applying phase advance to the altimeter signal direct may result in saturation of the missile servos. This can be avoided if the accelerometer and altimeter signals are mixed and filtered to produce the same effect; some ingenuity is called for here! Gravity compensation is easily provided. If the autopilot steady state gain is x m/sec^2/volt then an input bias voltage into the autopilot of g/x volts should result in a zero steady state height error in the absence of biases in the system. An alternative height lock loop can be designed using pitch gyro and altimeter feedback. Provided the missile speed is roughly constant a fixed input bias voltage can be used to result in just enough "nose up" body angle to give a normal force equivalent to 1g.

We now consider at what height the missile should be set to fly. A glance at "Janes Fighting Ships" (10) will convince one that a typical cruiser or aircraft carrier may well be an easy target but the small missile corvettes or fast attack craft such as the Russian Nanuchka or Osa class boats present a very low profile. Does one therefore fly relatively high with a very low probability of hitting a wave crest and then probably overfly the target, or does one fly lower and accept a small ditching probability? This question must be taken seriously especially as it is common knowledge that lethality increases the lower one hits a ship. Let us take a typical engagement with a range of 24 km, a missile velocity of 300 m/sec and a flight time of 80 secs. Firstly one must consider component biases. If an accelerometer is used and 50 m/sec^2 full scale deflection is adequate, a tolerance of 1% full scale (2σ value) is 0.5 m/sec^2. The rate gyro probably produces an equivalent bias. If one is thinking of a very low chance of ditching, say 0.5%, one could take a 2.8σ value which, assuming a normal distribution gives one a 99.5% chance of any one value being less than this. The total 2.8σ acceleration bias is therefore $1.4\sqrt{0.5^2 + 0.5^2} = 1$ m/sec^2. If the loop gain is 2 then the steady state height error is 1/2 = 0.5 m; this is because the loop gain or "stiffness" is 2 m/sec^2/m (height error). To this one must add a figure for the altimeter

error. If the 2σ value is taken as the tolerance, say 0.5m then the 2.8σ value is 0.7m. These two biases add to $\sqrt{0.5^2 + 0.7^2} = 0.86$ m. Now consider three particular sea states given by H_3rd = 1, 2 and 4m and assume for the moment that the sea state is approximately measured before the flight. If we take 2.8σ values for the sea state also and add say 25% to allow for the fact that the following error is likely to be greater than the sea "input" then a crude estimate of the mean height above sea level that a missile should be demanded to fly is shown in Table 6.14-1.

TABLE 6.14-1 ROUGH ESTIMATES OF "SAFE" HEIGHT TO FLY

H_3rd	1m	2m	4m
2.8 x 1.25σ	0.88	1.76	3.52
Instrument errors	0.86	0.86	0.86
Total following error	1.23	1.96	3.63
Add 0.4m for half wing span	1.63	2.36	4.03
Add 0.5m for incorrect estimation of sea state	2.13	2.86	4.53

The sea spectrum is obtained by weather ships by analysing the records of specially instrumented buoys. The question of estimating the sea state is therefore a difficult one although experienced sailors will say that they can judge the sea state accurately. Without experience however, estimates can be very unreliable; it is very difficult if the weapon is carried in a helicopter and the sea state has to be judged from the air. An arbitrary figure of 0.5m has been added in Table 6.14-1 to allow for this uncertainty.

An alternative is for the missile to estimate the sea state during the long run in period and to drop to a lower height during the last few seconds of the engagement. Consider now a perfectly smooth sea. The difference between the demanded height and the measured height must be zero if there are no biases in the autopilot. A bias in the altimeter will not result in a permanent error signal as this would cause the missile to have a constant vertical acceleration; this missile bias merely causes the missile to fly at the wrong height. If proportional plus integral control is used with a very long integrating time constant the height error signal will be reduced to zero and the stability of the loop will hardly be affected. In the presence of a real sea the error signal will be noisy but will have a zero mean. If the error is squared and then square rooted, the resulting signal if integrated over a period of 10 secs or more should be an accurate measure of the sea state. The height to fly during the last few seconds can be computed from a simple law thus

$$h_d = mx + c$$

where c allows for altimeter bias, wing size etc, x is the measured sea state and m is a constant.

6.15 VERTICAL LAUNCH

In chapter 3, some of the desirable aspects of vertical launch were discussed. Beam riding and CLOS systems will require a rapid and accurate turnover in order to enter the riding or tracker beam; thrust vector control would almost certainly be used to achieve this rapid turn during the missile boost phase. There are two distinct methods of implementing vertical launch. The first method consists of first rotating the missile about a vertical axis so that the missile oxz plane is parallel to the intended launch direction. A roll-pitch gyro and a yaw gyro can now be spun up and uncaged. To overcome the toppling problem the roll-pitch gyro spin axis is tipped back say 30^0 in the missile oxz plane. With pitch attitude feedback a simple turn in one plane is not difficult to programme.

A solution which is far simpler from the launcher point of view is to stack the missiles vertically as before but with no facility for rotating them about a vertical axis. Greater complexity however is required in the missile; one obvious method is to mount one or both of the gyros on a roll stabilised platform which is rotated by a servo before launch so that it is correctly orientated in the turnover plane. The feasibility of turning a missile over very rapidly in any direction relative to the initial missile orientation with the attendant large incidences would undoubtedly have to be proved by trials.

6.16 THE EFFECT OF FIN SERVO SATURATION

Since the input to an autopilot is invariably noisy it follows that the fin (or thrust vector) servos may easily become saturated. One obvious result of this is that the servos will not be able to implement true commands as accurately as they should. However, if saturation does not last long the system may well be in a position to recover completely. But saturation can have other less obvious effects. Consider a position servo with a sinusoidal input at a given amplitude and frequency. If the amplitude is small the servo will follow this input with a given amplitude ratio and phase lag. If now the amplitude or frequency is say doubled then the maximum angular velocity required at the output is doubled. Now the maximum velocity of any servo is limited: it is limited by the supply voltage in the case of an armature

controlled electric servo and it is limited by the maximum valve opening in a hydraulic or pneumatic servo. Limiting on velocity must result in a dynamic performance which is inferior to that which is defined by the nominal transfer function. A transfer function only defines a linear system and this amongst other things implies that saturation does not occur. Now the great number of autopilots rely on a very fast servo response to maintain loop stability. An increase in phase lag from the servo in the frequencies of interest will have a destabilising effect on the autopilot. "Rate limiting" as it is usually called would have to be really severe to unstabilise the autopilot completely. Nevertheless a significant reduction in autopilot stability implies a reduction in the effective damping and this means that the r.m.s. output of the autopilot due to noise is increased. This decreases the accuracy of the system and in addition decreases the effective range of the system due to the increased induced drag. Since noise will never consist of one amplitude and one frequency one can assume that what is important is the r.m.s. fin rate due to noise. If this r.m.s. rate approaches or exceeds the maximum velocity attainable by the servo then the system effectiveness is approaching one of the critical limits. This is the basic reason why missile servo specifications in general call for such a good dynamic performance. A design which calls for a high aerodynamic gain (e.g. zero static margin) and a low servo gain will help the servo designer in this respect.

We now consider the likely effect angle limiting will have on the servo performance since all fin servos or thrust vector servos are constructed with mechanical stops to limit the angular travel. Angle limiting will clearly have the effect of reducing the r.m.s. gain of the servo. If the servo is not limiting on rate it means that the time (or frequency) response must be improved as the servo will not have to work so hard had there been no limits. This will usually show as a small reduction in phase lag at any given frequency.

Finally, it should be pointed out that the precise simulation of limiting is difficult (e.g. does the fin velocity fall to zero when a mechanical stop is reached, or does the load bounce?). An investigation of the effect of the servo dynamics on the autopilot performance is best done with the actual servo and simulated aerodynamics and instrument feedback. Such simulations that have been carried out by the author suggest that in practice rate limiting is more likely to have a significant effect on the autopilot performance and since the effect is invariably bad one concludes that a realistic specification for the fin servos is essential. It is very easy to over-specify so that one

is on the safe side but this often results in a lot of extra development work, instead of being able to use a conventional or existing design.

REFERENCES

1. GRAHAM D. and LATHROP R.C. 1953 Trans. Am. Inst. Elect. Engrs, 72, 273.

2. TOWILL D.R. Transfer function techniques for control engineers. Iliffe Books Ltd 1970.

3. DARWELL H.M. and TRUBRIDGE G.F.P. Design of rocket nozzles to reduce gas misalignment. Journal of Spacecraft and Rockets. January 1968.

4. International Defence Review. April 1971, pp 168-9.

5. FRARY D.J. The prediction of autopilot behaviour in the presence of roll motion. British Aircraft Corporation Report no. ST 5686. May 1971.

6. DARBYSHIRE J. Proc. Royal Soc. A.215 Vol 215 1952 No. 1112 pp 299-328 "The generation of waves by wind".

7. HOGBEN N. and LAMB F.E. Ocean Wave Statistics. H.M. Stationery Office 1967.

8. NEVILLE E.J. Standard wave spectra. National Physical Laboratory Ships TM 301 March 1971.

9. MURPHY C.H. Free flight motion of symmetric missiles. Aberdeen Proving Ground report no. 1216 July 1963.

10. Jane's Fighting Ships 1975-76. Jane's Yearbooks.

CHAPTER 7

LINE OF SIGHT GUIDANCE LOOPS

7.1 THE EFFECT OF TARGET AND MISSILE MOTION ON MISSILE "G" REQUIREMENTS

There are many types of LOS systems but before examining some of the differences between them we shall consider some of the common aspects. LOS systems can be called "3 point guidance" systems since there is one point which defines the tracker, another the target and a third which defines the position of the missile. The object of the guidance system is to constrain the missile to lie as nearly as possible on the line joining the tracker and the target called the line of sight (LOS). The concept is simple and can be implemented in many ways; perhaps it is this apparent simplicity which explains why the majority of all guided weapon systems as yet designed are LOS systems. Nevertheless they all exhibit some fundamental characteristics which limit their performance and coverage.

Consider a target flying straight and at constant speed, and a missile flying at a different but constant speed, having been launched when the target occupies a position T_0, see Fig 7.1-1. After intervals of time of 1, 2, 3 etc seconds

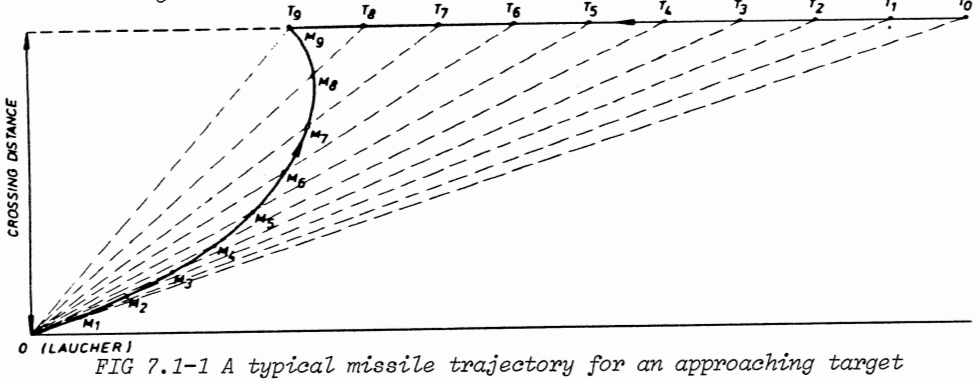


FIG 7.1-1 A typical missile trajectory for an approaching target

the LOS is shown as $0T_1$, $0T_2$, $0T_3$ etc. Since the missile ideally always lies on these lines the flight path will be a curved one and, for an approaching target, the curvature becomes increasingly severe towards the end of the engagement. We note that the tangent to the flight path at any one point defines the instantaneous direction of the missile velocity. It is seen that the missile velocity vector will, in general, not be directed along the LOS; towards the end of the engagement it may be at a considerable angle to it.

Since a LOS is often referred to as a beam, let us call this angle "the trajectory-to-beam angle". However, what is really significant is the body-to-beam angle and this will exceed the trajectory-to-beam angle by the angle of incidence (see Fig 4.4-1). The reader can check this by noting from Fig 7.1-1 that the lateral acceleration must be to the left looking down the missile flight path; this means that the relative wind must be coming on to the missile's right side and therefore the front end of the missile must be to the left of the flight path.

In some LOS systems the missile carries a beacon to transmit position information to the ground tracker. The beacon will have a limited beam-width and if the body-to-beam angle exceeds this a serious loss of signal will result. If one tracks the missile with the aid of retroreflectors the same argument applies since their effectiveness falls off with increasing angle. A 40^{o} limit is often quoted as reasonable for beacons, 45^{o} for retroreflectors, and 50^{o} for beam-riding and command link receivers. These figures are only typical ones but whatever value is agreed upon for a given system, a constraint on the missile coverage will automatically follow. Body-to-beam angle is also an important parameter since the missile has to develop a component of acceleration perpendicular to the beam due to the beam motion and as the body-to-beam angle approaches 90^{o} so does the aerodynamic manoeuvring force developed perpendicular to the missile body become less and less effective, thus leading to larger and larger demands for lateral acceleration ("latax").

If now a body is moving along a rotating line at a velocity U then it will experience an acceleration perpendicular to the line equal to the Coriolis component $(2U\dot{\theta})$ + a component equal to $R\ddot{\theta}$ where R is the distance from the stationary centre of rotation. When the velocity vector U_m is at an angle σ_m to this line and the lateral acceleration is assumed to be developed perpendicular to the velocity vector we find that the Coriolis component is unchanged because the component of velocity along the line (the LOS in our case) is reduced by a factor equal to cos σ_m but the useful component of latax (i.e. that component perpendicular to the LOS) is proportional to $1/\cos\ \sigma_m$. However the latax the missile has to pull due to the $\ddot{\theta}$ term is increased by a factor equal to $1/\cos\ \sigma_m$. If in addition the missile has a component of acceleration $\dot{U}_m$ in the direction of the velocity vector (i.e. it is accelerating or decelerating in the conventional sense) there will be a component of acceleration perpendicular to the beam of $\dot{U}_m \sin\ \sigma_m$. Hence we can write down the latax required perpendicular to the velocity vector

for a missile at a range R_m from the tracker due to a target at R_t from the tracker thus:

$$f = 2U_m\dot{\theta} + R_m\ddot{\theta}/\cos\sigma_m - \dot{U}_m \sin\sigma_m/\cos\sigma_m$$
$$= 2U_m\dot{\theta} + R_m\ddot{\theta}/\cos\sigma_m - \dot{U}_m \tan\sigma_m \qquad (7.1\text{-}1)$$

Now $\dot{\theta}$ and $\ddot{\theta}$ clearly depend on the target velocity and position. If reference is made to Fig 7.1-2 we can write

$$\dot{\theta} = U_m \sin\sigma_m/R_m = U_t \sin\sigma_t/R_t$$

$$\text{i.e. } \sin\sigma_m = \sin\sigma_t \frac{R_m}{R_t}\frac{U_t}{U_m} \qquad (7.1\text{-}2)$$

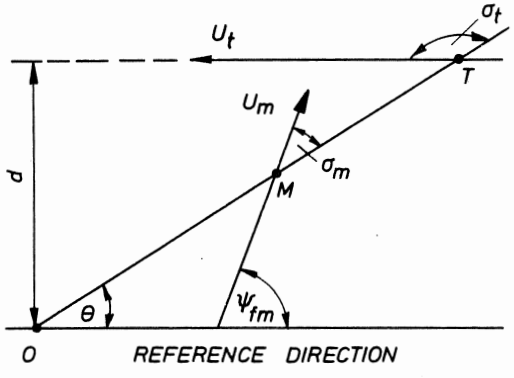


FIG 7.1-2 Flyplane geometry

For a straight flying target the crossing distance d is a constant so another expression for $\dot{\theta}$ is given by

$$\dot{\theta} = U_t \sin^2\sigma_t/d \qquad (7.1\text{-}3)$$

and $\ddot{\theta}$ is obtained by recalling that $\ddot{\theta} = -\dot{\sigma}_t$ and

$$\frac{d^2\theta}{dt^2} = \frac{d}{d\theta}\left(\frac{d\theta}{dt}\right).\left(\frac{d\theta}{dt}\right) = -\frac{2U_t^2}{d^2}\sin^3\sigma_t\cos\sigma_t \qquad (7.1\text{-}4)$$

Substituting these expressions for $\dot{\theta}$ and $\ddot{\theta}$ in equation (7.1-1) and rearranging for a constant speed missile yields

$$f = \frac{2U_mU_t}{d}\frac{\sin\sigma_t \sin(\sigma_t - \sigma_m)}{\cos\sigma_m} = \frac{2U_mU_t}{d} \times \alpha \qquad (7.1\text{-}5)$$

where α is a non-dimensional factor and is a function of σ_t and σ_m. But if σ_t is given then σ_m can be calculated using equation (7.1-2). Fig 7.1-3 shows the relationship between σ_m and σ_t for discreet values of $R_m\,U_t/R_t\,U_m$ and overlaid are lines of constant acceleration factor α.

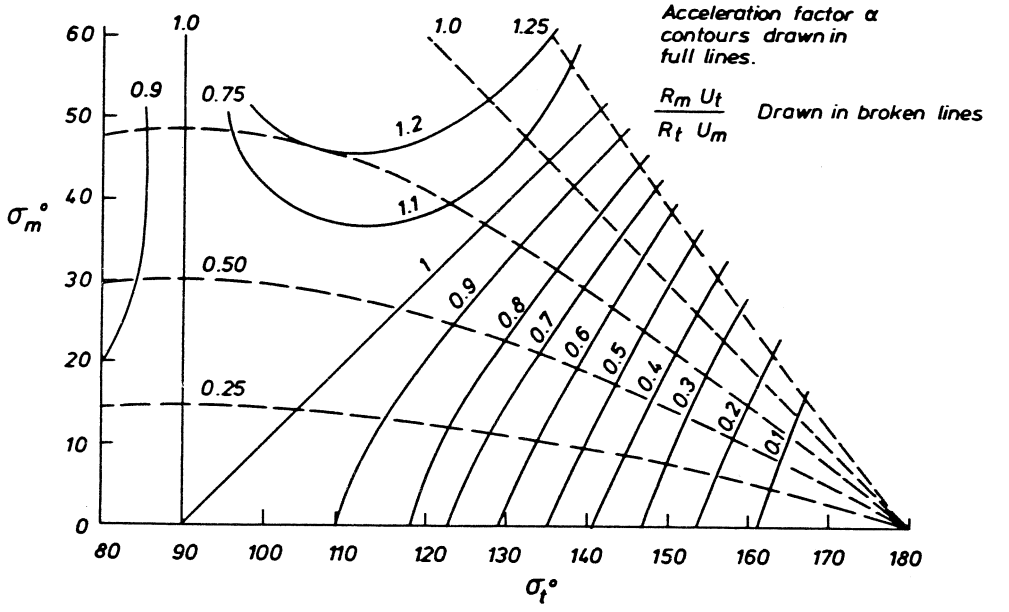


FIG 7.1-3 Trajectory-to-beam angles and acceleration factors as a function of σ_t

If $\dot{U}_m$ is known then an additional component as already noted in equation (7.1-1) is involved. If the missile is coasting and therefore decelerating, the longitudinal acceleration is say 60 m/sec^2 and σ_m is 50^0 then the additional latax the missile has to pull is given by 60 tan 50^0 = 71.5 m/sec^2. This is the one great disadvantage of a boost-coast velocity profile; missile g requirements are, apparently, increased.

What general conclusions can we draw from Fig 7.1-3? Firstly it is obvious that latax requirements are greatest in the region of σ_t equal to 130^0 to 100^0 and this is for values of θ between 50^0 and 80^0. A speed superiority is an advantage as this helps to keep trajectory-to-beam angles to reasonable values. At crossing point the angular acceleration is zero and beyond crossing point ($\sigma_t{}^0 < 90^0$) the angular acceleration is negative. Equation (7.1-1) therefore indicates that lateral acceleration requirements tend to be less for receding targets. Does equation (7.1-5) imply that if the crossing distance d is very small the latax requirements are always very large? Consider the limiting case of d = 0, i.e. a target flying straight at the tracker. In such a case the sightline has zero rate and zero angular acceleration and σ_m = 0, and therefore the missile ideally flies straight with zero latax. Equation (7.1-5) reduces to

$$f = \frac{2U_m U_t}{0} \sin^2 0^o = 0$$

If now we think of a target at a given slant range then a reduction in crossing distance will result in a reduction in θ i.e. σ_t tends towards 180^o. Therefore short crossing distances for a given range of target result in small latax requirements as we are implying that the target is more or less heading straight to the tracker. The really difficult task therefore is area defence and not point defence. Even so, the latax requirements will be moderate or low for an area defence system if the impact range exceeds 2 or 3 km say since $\dot{\theta}$ is inversely proportional to R_t and $\ddot{\theta}$ is inversely proportional to R_t^2 all other things being equal. Table 7.1-1 shows the lateral acceleration required at interception for U_t = 270 m/sec and U_m (assumed constant) = 540 m/sec for different crossing distances and interception ranges R_{ti}.

TABLE 7.1-1 LATERAL ACCELERATION REQUIRED AT INTERCEPTION

R_{ti} \ d	100m	200m	500m	2 km
0.5 km	173	342	583	Geometry not possible
2 km	10.9	21.8	54.4	146
3 km	4.8	9.6	24.2	90.4

Short range missiles can be designed to pull more than 20g if the speed is in excess of about M = 1.5 and even 30g for M = 2 say, but only moderate latax is feasible if the speed falls to M = 1 or less. The latax a missile can pull in practice is limited not only by the dynamic pressure and the wing etc size but by the maximum incidence it is feasible to allow; and this leads us to a discussion on the difference between the results shown in Fig 7.1-3 and Table 7.1-1 and those which would obtain if the missile incidence is taken into account, as indeed it must be in practice. Consider the case when the trajectory-to-beam angle is 10^o and the incidence is say 5^o. Since the cosines of 10^o and 15^o are still nearly unity the $\dot{\theta}$ and $\ddot{\theta}$ terms are both increased by a negligible factor of cos 10^o/cos 15^o and the $\dot{U}$ term is increased by a factor of tan 15^o/tan 10^o. In other words when the angles are small the effect of incidence is practically negligible. This is not so when trajectory-to-beam angles are 40^o, especially when one notes that incidence is likely to be greatest when these angles are large. Even ignoring the beacon/retroreflector/receiver problem then, one sees that lateral acceleration requirements can escalate to ridiculous values if body-to-beam angles are large especially if the missile is coasting.

What then is the ideal velocity profile for a missile? Clearly for a manual system when the human operator tracks both the target and the missile, a constant velocity missile makes it possible to design a missile with a near-consistent response without the use of instrument feedback. With a short range air-to-air system when flight times could be a few seconds only and targets may be approaching, crossing or receding, a motor which gives a moderate boost for about 10g for 4-5 seconds has the advantage that $\dot{U}_m$ is positive during the critical period when latax requirements can be exceptionally high. A missile which is launched from an aircraft starts with a fair speed increment but if the initial manoeuvrability is not sufficient a system which uses thrust vector control and is boosting offers exceptional manoeuvrability. There are strong arguments for a boost-coast propulsion motor giving say 40g for 2 seconds for a surface-to-air missile designed to hit approaching targets. We have seen that the highest latax demands tend to occur at short ranges so, if due to late acquisition of the target or due to the sheer speed of the oncoming target the impact range is short we require maximum manoeuvrability of the missile as soon as possible, and this requires a high missile speed. If the target speed is low or acquisition is early impact will occur at a greater range and therefore latax requirements will be less; a lower missile speed at greater range will probably not restrict the coverage. Also, a boost-coast requirement results in a simple motor whose nozzle exit area can be designed for maximum thrust. A motor which has to thrust at two levels is more complex and will be less efficient as the nozzle design must be a compromise. Also a boost-coast velocity profile enables the missile, in the first 2 or 3 seconds, to get to a greater range than one with a lower boost followed by a sustain phase and this can be vitally important in simple systems where the surveillance is such that it is improbable that targets will be first sighted at ranges in excess of 4-5 km. This tends to reduce latax requirements. Finally if at any time the operator can be involved in tracking, the absence of smoke in the coast phase may facilitate his tracking task.

A useful indication of the kinematic conditions governing coverage for a boost-coast missile is given by Gridley (1). It is assumed that the missile is boosted instantaneously to a speed B times that of the velocity of the target whose speed is assumed to be constant. It is also assumed that the deceleration of the missile is proportional to velocity, a reasonable approximation for supersonic conditions i.e.

$$\dot{U}_m = - U_m/\tau$$

where τ typically has a value of 8-10 for many supersonic missiles. If one assumes that the manoeuvring force is developed perpendicular to the velocity vector, solutions for the latax required and the trajectory-to-beam angles *at interception* can be obtained for a particular value of B. The following non-dimensional parameters are defined

$$r = R/U_t\tau \text{ where } R \text{ is the range at interception}$$

$$\text{and } A = f\tau/U_t$$

The graphs shown in Figs 7.1-4(a), (b) and (c) approximate though they may be on account of the assumptions made, do indicate the very small coverage provided by a missile against a fast target unless it is boosted to at least twice the target speed. For example, if τ = 10 and U_t = 200 then the maximum crossing distance is approximately given by

$$r = 0.5 \text{ for } B = 1.5, \text{ i.e. } d = 1000\text{m}$$

$$\text{and } r = 1.0 \text{ for } B = 2.0, \text{ i.e. } d = 2000\text{m}$$

Again, it is emphasised that as σ_m increases, so the graphs must become increasingly inaccurate on account of the body incidence becoming reckonable.

7.2 TYPES OF LOS SYSTEMS

There are many types of LOS systems but the great majority fall into one of the four main types broadly described below.

Beam riding systems

Beam riding systems use a target tracker whose purpose is to maintain the antenna boresight pointing at the centre of the reflecting area of the target. We have discussed some aspects of target trackers in chapter 1. The missile carries a complicated receiver which can detect the missile's angular deviation from the centre of the beam. The steering signals designed to bring the missile to the centre of the beam are therefore generated in the missile itself and a missile tracker is not required. To keep the sensitivity (volts/m off the beam) of the receiver constant its output is effectively multiplied by a factor proportional to the range (usually assumed) from the launcher, as already discussed in chapter 6.1. Early missiles achieved this by feeding one end of a potentiometer with the receiver output. A slider is driven by an electric motor through a speed reducer and the voltage at the slider is the modified error signal. If the speed of the missile is approximately constant a linear potentiometer and constant speed motor suffices. If the missile speed is not constant, a specially wound non-linear potentiometer can be designed. Today very accurate integrated circuit multipliers are available

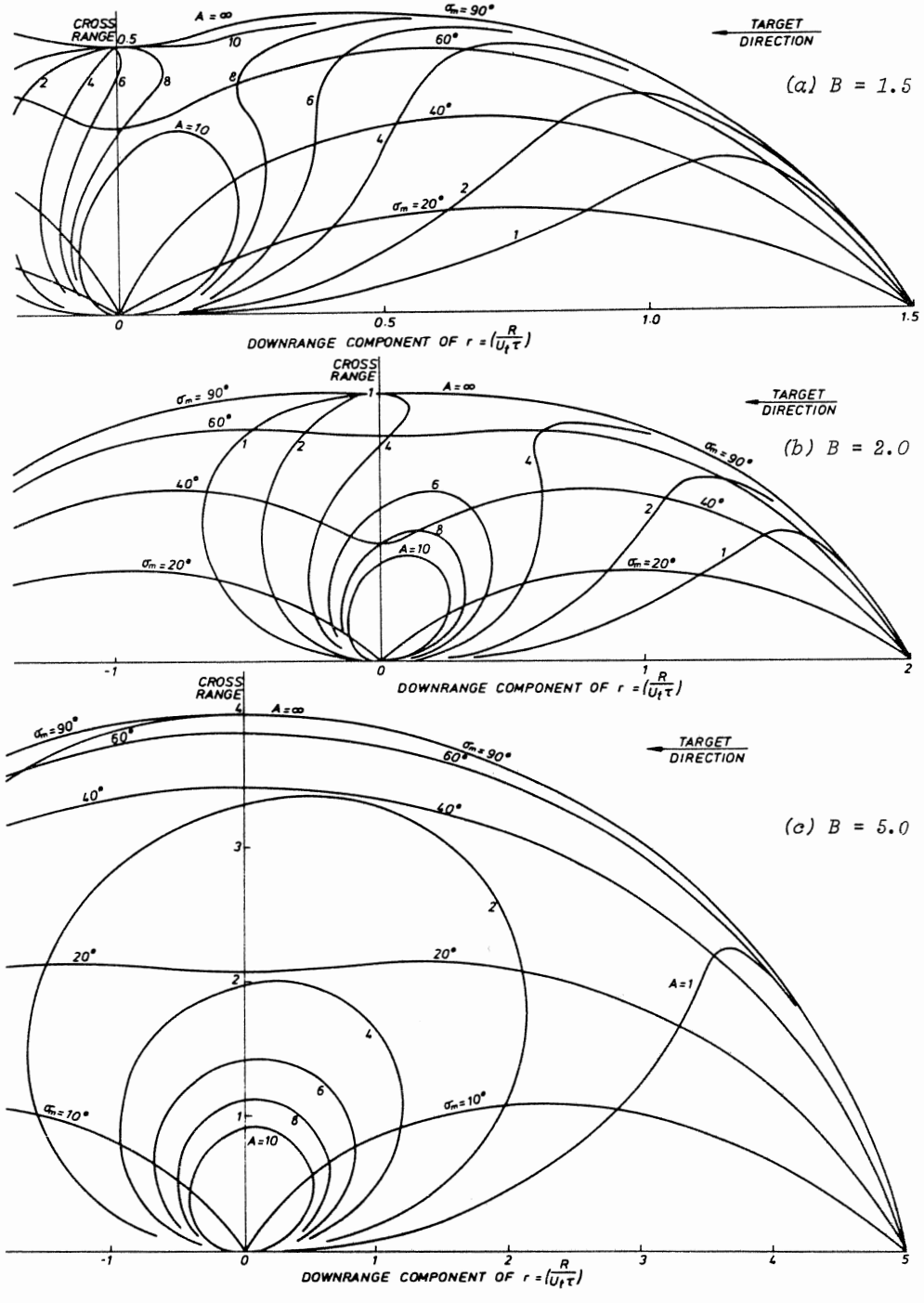


FIG 7.1-4 Acceleration contours and trajectory-to-beam angles for a boost-coast missile for different values of B

one input to which can be a generated voltage proportional to the assumed nominal range, and the other input the signal from the receiver. Fig 7.2-1 shows in block diagram form, the general structure in one plane of a beam-riding system.

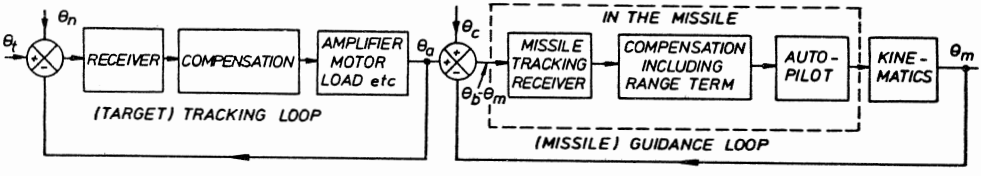


FIG 7.2-1 Beam-riding system-tracking and guidance loops

If the system uses one beam for tracking and another beam for guiding the missile there will be the problem of keeping these two beams truly parallel and a collimation error θ_c has been shown as an input to the guidance loop. If the same beam is used for tracking the target and guiding the missile the question of collimation error does not arise.

Manual systems

The expression "a manual guided weapon system" implies that the human operator has the task of tracking both the target and the missile, and generating commands to bring the missile into the LOS. Most, but not all manual systems that have been designed have been anti-tank systems. For ranges in excess of about 800m some target magnification is required, so the operator has to keep the target within the field of view of the sight and generate commands with the aid of a specially designed thumb controller. It should be particularly noted that the operator is acting as a differential tracker; provided he can keep both target and missile in the field of view the absolute tracking accuracy is not relevant. With help from training simulators the human operator can develop into a highly efficient guided missile controller; his task is always regarded as a little easier when some defined background is present as the results of his commands are easily discernible. This is usually referred to as a "pursuit" tracking task. In a surface-to-air situation with a clear sky, he is only aware of the angular error and the results of his commands are not so easy to interpret especially if the target is manoeuvring; such a task is usually referred to as "compensatory" tracking. Despite the obvious simplicity of employing an operator to perform the tracking task a manual system is rarely considered now as a serious competitor to semi-automatic systems. Unfortunately it does take time to train an operator even with the aid of a good simulator and hence the weapon must be regarded as a specialist

weapon. Also, bad weather can greatly reduce his effectiveness and there is always a limitation on the minimum effective range since it will probably take the operator at least 3 seconds after firing before he has the missile completely under control. An automatic or semi-automatic system should do better than this.

Semi-automatic systems

In a semi-automatic system the human operator only tracks the target; his task is to keep the optical centre line of his sight pointing at the required point of missile impact. His sight will be fitted with cross-wires or an etched spot on a graticule to assist him in aiming, and his sight may well be servo driven in azimuth and elevation as a result of commands generated from a thumb or thumb and finger held joystick. Fig 7.2-2 indicates some of the well known features of the surface-to-air guided weapon system Rapier (2).

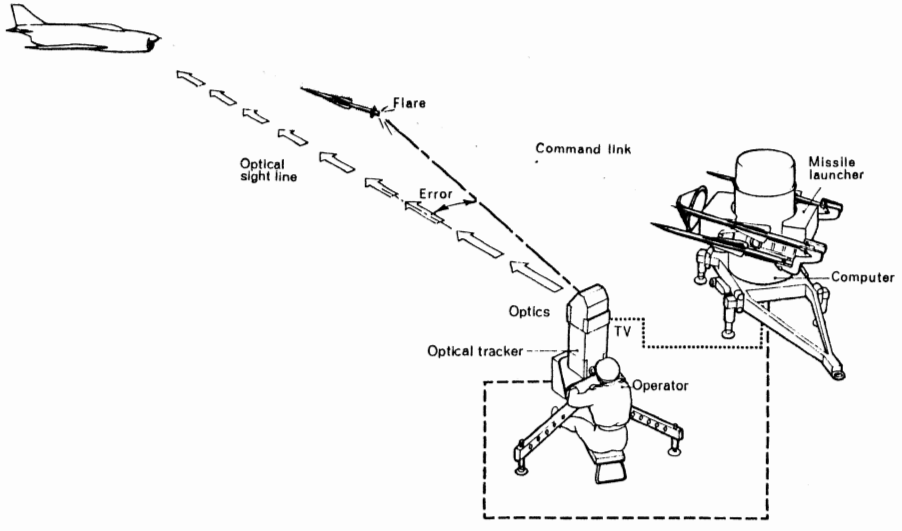


FIG 7.2-2 RAPIER - a semi-automatic CLOS system

The missile tracker is mounted alongside the operator's tracking telescope and is collimated to it. Signals proportional to the angular misalignment of the missile from the missile tracker boresight are processed by the ground computer and transmitted to the missile by means of a radio link; we have already seen that "processing" will include multiplication by a term proportional to the missile range, and this is easily achieved if a digital computer is available. In other words, the guidance loop is essentially no different from a beam-riding guidance loop.

However, most semi-automatic and automatic LOS systems except beam riders use feed-forward terms in order to increase accuracy; this special feature will be considered in detail later in the chapter. Fig 7.2-3 shows the general

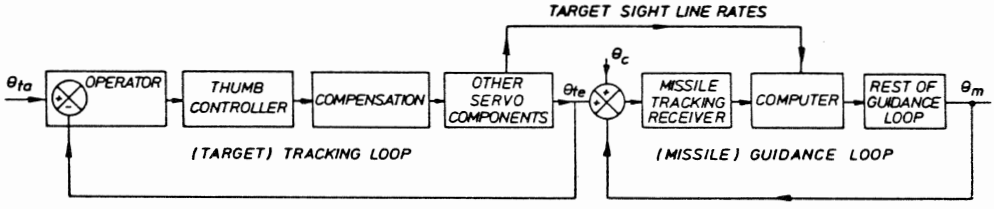


FIG 7.2-3 Semi-automatic system - tracking and guidance loops

structure of a semi-automatic guidance and control system using feed-forward terms. Sight line rates must be measured in such a system. Since commands are computed on the ground there must be a command link between the computer and the missile; this is often a wire link for anti-tank systems but is invariably a radio link for systems where the missile is supersonic. Since there are two completely separate tracking systems there must be some collimation error or disturbance θ_c.

LOS systems using differential trackers

A differential tracker is designed to eliminate the inaccuracy due to imperfect target tracking. A simplified diagram of a differential tracker is shown in Fig 7.2-4. The antenna is servo controlled to follow the target direction in

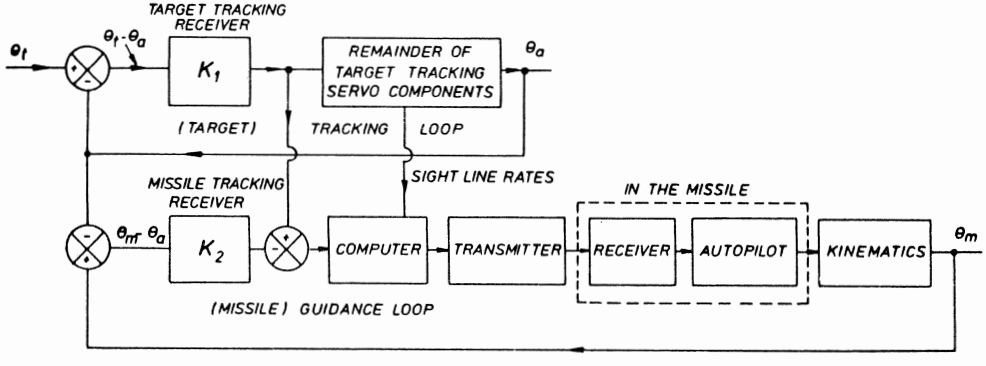


FIG 7.2-4 Differential tracking system - tracking and guidance loops

the usual way. By multiple beam forming or by multiplexing, the same aperture is used to obtain a signal proportional to the misalignment between the antenna boresight and the missile direction. This signal is subtracted from the target tracker error signal. This gives us a signal equal to

$$(\theta_t - \theta_a)k_1 - (\theta_m - \theta_a)k_2$$

and this reduces to $(\theta_t - \theta_m)k_1$ if $k_1 = k_2$ i.e. the antenna direction is eliminated. However, even if the same receiver is used to make these two measurements, it is not easy to maintain the gains of these two measurements identical, and special feedback and calibration techniques are used. It should be particularly noted that in all other systems the target tracking loop acts as a low-pass filter to thermal noise and glint. There is no such filter outside the guidance loop with a differential tracker. Since filtering cannot be introduced inside the missile guidance loop without reducing stability margins one must conclude that with differential trackers the effect of noise associated with tracking the target will have a greater effect on miss distance. It is just possible with a great deal of ingenuity to introduce a small amount of filtering of high frequency noise outside the guidance loop.

7.3 KINEMATIC CLOSURE AND STABILITY OF THE GUIDANCE LOOP

If a missile produces a lateral acceleration as a result of a guidance command, what is the kinematic relationship between this acceleration and the change of angle the missile presents to the missile tracker which we assume to be stationary? Equation (7.1-1) gives the lateral acceleration required on the assumption that it is developed perpendicular to the velocity vector; but we know that lateral forces are developed perpendicular to the body. Considering the yaw plane only and assuming the body incidence is β we see that the useful component of lateral acceleration is the component perpendicular to the beam as the component along the beam in no way helps to reduce an angular error. If f_y is the missile latax then the useful component is $f_y \cos(\sigma_m + \beta)$. A serious loss in guidance loop gain will occur if the body-to-beam angle exceeds say 45^o, and this is a very good reason for avoiding large body-to-beam angles.

Up till now we have made the assumption that the useful missile latax divided by the missile range and integrated twice is the change in missile angle as seen by the angle channel receiver. This assumption is reasonable for a stability analysis but is nevertheless not strictly accurate when simulating a guidance loop: the complication is due to the missile range not being constant. Referring to Fig 7.1-2 we can write $R_m\dot{\theta} = U_m \sin \sigma_m$ and this can be differentiated to read

$$\dot{R}_m\dot{\theta} + R_m\ddot{\theta} = U_m \cos\sigma_m \dot{\sigma}_m \text{ (for } U_m \text{ a constant)}$$
$$= U_m \cos\sigma_m (\dot{\psi}_{fm} - \dot{\theta}) \qquad (7.3\text{-}1)$$

and since $\dot{R}_m = U_m \cos\sigma_m$ we can substitute in equation (7.3-1) and rearrange to read

$$(TD + 1)\dot{\theta} = \dot{\psi}_{fm}/2 \qquad (7.3\text{-}2)$$
$$\text{where } T = R_m/2U_m \cos\sigma_m$$

This time constant is therefore very roughly one half of the time from launch if one makes the approximation that $R_m = U_m\tau$ where τ is the time from launch and $\cos\sigma_m$ is not far from unity. Equation (7.3-2) can be arranged to read

$$\left(D + \frac{2U_m \cos\sigma_m}{R_m}\right) R_m\dot{\theta} = U_m \cos\sigma_m \dot{\psi}_{fm} = f_y \cos\sigma_m$$

$$\text{since } f_y = U_m \dot{\psi}_{fm}$$

Hence, in transfer function form

$$\frac{\theta}{f_y} = \frac{\cos\sigma_m}{s R_m (s + 2U_m \cos\sigma_m/R_m)} \qquad (7.3\text{-}3)$$

and this contains one integration and a time - varying time constant. For stability purposes we now have to demonstrate that this time constant approximates to pure integration in the critical frequency range. If the guidance open loop gain crossover frequency is 4 rad/sec, $U_m = 500$, $R_m = 5000 \cos\sigma_m \simeq 1$ then at this frequency this time constant is equivalent, in pure frequency form to

$$\frac{1}{\sqrt{4j} + 1/5} \equiv \frac{1}{4.008} \angle -87^o$$

instead of $\frac{1}{4j}$.

It is considered therefore that since the guidance loop has to be stable at all times and ranges, it is a perfectly realistic assumption to regard the kinematic closure of the guidance loop from lateral acceleration to position to result in 180^o phase lag. We now consider the most desirable form of the compensation necessary to ensure an adequate stability margin in the guidance loop. It has already been shown in chapter 1 that if a closed loop system contains two integrations for a given open loop gain the minimum noise bandwidth is obtained if the closed loop is a second order system with a damping ratio of 0.5; and this is on the assumption that the compensation is in the feed-forward path. Additional lags increase the equivalent noise bandwidth. Now a damping ratio of 0.5, for a second order system requires a phase margin of 52^o. Since the autopilot must exhibit some phase lag at gain crossover frequency we are really looking for at least 60^o phase advance.

Suppose now we (quite arbitrarily) limit the amplification of high frequency noise to a figure of 10 to 1. A single phase advance circuit gives a maximum phase advance of 55^{o} (see equation 6.2-1). A double phase network of the form

$$\left(\frac{1 + Ts}{1 + \alpha Ts}\right)\left(\frac{1 + Ts}{1 + \alpha Ts}\right)$$

where $\alpha = \sqrt{0.1}$ gives a maximum phase advance of 62.6^{o}. Indeed a treble phase advance network where $\alpha = \sqrt[3]{0.1}$ will give a maximum phase advance of 64.4^{o} but we are now in a region of rapidly decreasing returns. Fig 7.3-1 shows the gain and phase characteristics of a 10 to 1 single active phase advance network

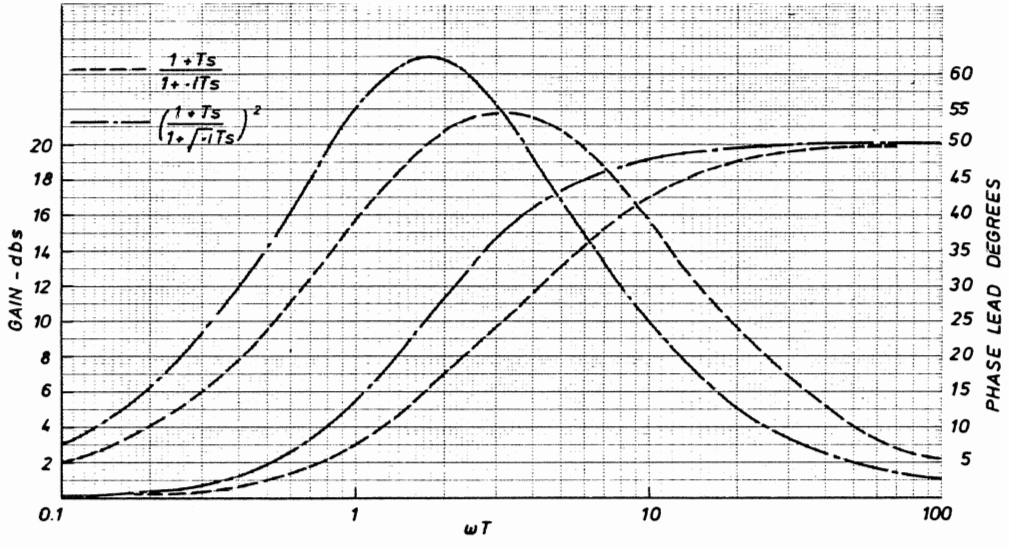


FIG 7.3-1 Gain and phase characteristics of phase advance networks

compared with that of two cascaded $\sqrt{10}$ to 1 networks. The phase characteristics of the cascaded arrangement are more "peaky" than that of the single network. Those of a treble network become even more peaky and therefore such an arrangement is seriously detuned if the gain crossover frequency should change significantly due to a large body-to-beam angle which reduces the gain of the loop. There appears to be no real advantage in going for a treble phase advance network.

It is always useful to check the equivalent noise bandwidth of any proposed guidance loop. With two integrations in the loop open loop gain crossover frequency would occur at $\omega_c = \sqrt{K}$ where K is the guidance loop d.c. gain or stiffness. Since the gain is decreasing at 40 db/decade and a 10 to 1 phase advance network has a gain of 10 db at maximum phase advance the new gain

crossover frequency will be $\sqrt[4]{10}\,\omega_c$ if unity gain is to coincide with maximum phase margin; this argument holds for a cascaded arrangement also. Since maximum phase advance occurs when $\omega T = 1/\sqrt{\alpha}$ the reader will find that the optimum value of T is given by $T = 1.778/\sqrt{K}$ for a single network, and $T = 1/\sqrt{K}$ for a double phase advance network.

Inserting these values into a system with two integrators and evaluating the noise bandwidths as given in Table 1.4-1 we find that it is

$$\frac{\pi}{2}\sqrt{K} \;.\; 2.61 \text{ for a single phase advance network}$$

$$\text{and } \frac{\pi}{2}\sqrt{K} \;.\; 2.51 \text{ for a double phase advance network}$$

compared with $\frac{\pi}{2}\sqrt{K}$. 2 for a simple second order system with stabilisation in the feedforward path and $\mu = 0.5$.

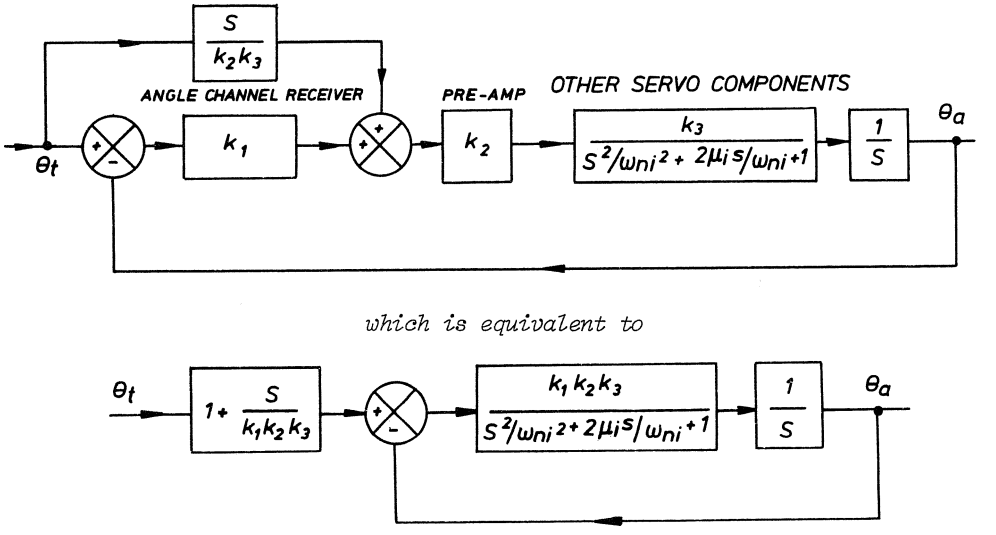
In practice the double phase advance network will probably compare rather more favourably than indicated above since there will certainly be some phase lag due to the autopilot (10^0 at gain crossover frequency?) and possibly a small phase loss if a digital computer is incorporated into the guidance loop. Also, it may be worth while inserting a small filter - an exponential lag of a few milliseconds - just before the demand to the autopilot to filter out some of the relatively high frequency noise. Our stability margin is, in practice, always less than the optimum and a few extra degrees of phase margin obtained by the use of a cascaded network will nearly always be worth having.

7.4 THE CONCEPT OF FEEDFORWARD TERMS

The term "feedforward" needs some redefinition at this stage. We are all aware of the fact that any closed loop system consists of a feedforward path (containing feedforward elements) and a feedback path. If integral control is used we say that we "feedforward the integral of the error", meaning of course that this is in addition to the proportional controlling signal. In the same way we can feedforward the differential of the error. But there is another method available of increasing the accuracy of a closed loop system and that is feedforward of some function of the *input*. Suppose we wish to eliminate velocity lag in the tracking servo described in chapter 1 without using integral control. Without this extra integrator the system transfer function is

$$\frac{\theta_a}{\theta_t} = \frac{1}{s^3/\omega_{ni}^2\, k_1k_2k_3 + 2\mu_1 s^2/\omega_{ni}\, k_1k_2k_3 + s/k_1k_2k_3 + 1} \qquad (7.4\text{-}1)$$

The voltage from the receiver necessary to drive the antenna at unit speed is $1/k_2k_3$ and hence if it were possible to generate *external to the closed loop* a voltage proportional to the target angular rate and feed it into the feed-forward path as shown in Fig 7.4-1 all errors due to the input angular rate would be eliminated. However, this method is essentially an open-loop method



which is equivalent to

FIG 7.4-1 Feedforward of derivative of input

of compensation and relies on being able to match this additional input to some servo component gains. The system transfer function is now

$$\frac{\theta_a}{\theta_t} = \frac{s/k_1k_2k_3 + 1}{s^3/\omega_{ni}^2\, k_1k_2k_3 + 2\mu_i s^2/\omega_{ni}\, k_1k_2k_3 + s/k_1k_2k_3 + 1} \qquad (7.4\text{-}2)$$

and it is seen that the coefficients of "s" in the numerator and denominator are nominally equal, which is the condition for zero velocity error. As it so happens this technique cannot be applied to a tracking radar as the only physical quantity available is the angle channel receiver output. In the case of a target tracker acting as a director to a gun or launcher the inputs to the latter are the angular positions of the tracker azimuth and elevation shafts. Tachogenerators can be attached to these "inputs" to provide a signal proportional to input rate. The use of input feedforward is common in such systems.

The use of additional feedforward terms in guidance loops is precisely the same in principle except that one tries to be even more precise; one tries to eliminate the error whatever the nature of the input. Simple feedforward of

some function of the input just will not do in a guidance loop. This is due to the fact that there is a x R_m term explicitly associated with the guidance functions and a $1/R_m$ term associated with the kinematic closure. This leads to the curious result that a lateral acceleration is required even if the input is a constant angular velocity. To regard a guidance loop as a straightforward type 2 closed loop would be quite wrong. So we adopt a different approach and calculate on the ground as a continuous function of time precisely what demand is required to keep the missile in the LOS; assuming the missile produces a lateral acceleration as a result of a command this means that we require to know the missile latax required. It would be convenient to arrive at an expression for missile latax which involves terms we can either measure easily or can assume with some degree of confidence. Firstly we will calculate the latax required due to target motion in the flyplane. It is clear that whether the target flies straight or manoeuvres in any manner it will remain in the flyplane provided there is no target acceleration perpendicular to the flyplane. Hence the missile velocity vector must always lie in the flyplane and no latax perpendicular to the flyplane is required. Denote the missile latax in the flyplane by f_1. Now the lateral acceleration is the product of the missile forward velocity and the rate of change of flight path (see 4.4-3) and if reference is made to Fig 7.1-2 it is seen that

$$f_1 = U_m \dot{\psi}_{fm} = U_m (\dot{\theta} + \dot{\sigma}_m) \qquad (7.4\text{-}3)$$

In general, the flyplane will be neither vertical nor horizontal, see Fig 7.4-2. If tachogenerators are fitted to the tracker azimuth and elevation shafts, to measure $\dot{\gamma}$ and $\dot{\phi}$ respectively then total angular rate $\dot{\theta}$ can be computed from

$$\dot{\theta} = \sqrt{\dot{\gamma}^2 \cos^2 \phi + \dot{\phi}^2} \qquad (7.4\text{-}4)$$

There is now a computing problem since there is no method of measuring σ_m or $\dot{\sigma}_m$. In equation 7.1-2 we used the relationship that $\sin \sigma_m = \dot{\theta} R_m / U_m$ and this can be differentiated to read

$$\frac{d}{dt} \sin \sigma_m = \frac{d}{d\sigma_m} \sin \sigma_m \frac{d\sigma_m}{dt} = \cos \sigma_m \frac{d\sigma_m}{dt}$$

$$= \dot{\theta} \frac{d}{dt} \left(\frac{R_m}{U_m}\right) + \left(\frac{R_m}{U_m}\right) \ddot{\theta}$$

$$\text{But } \cos \sigma_m = \sqrt{1 - \sin^2 \sigma_m}$$

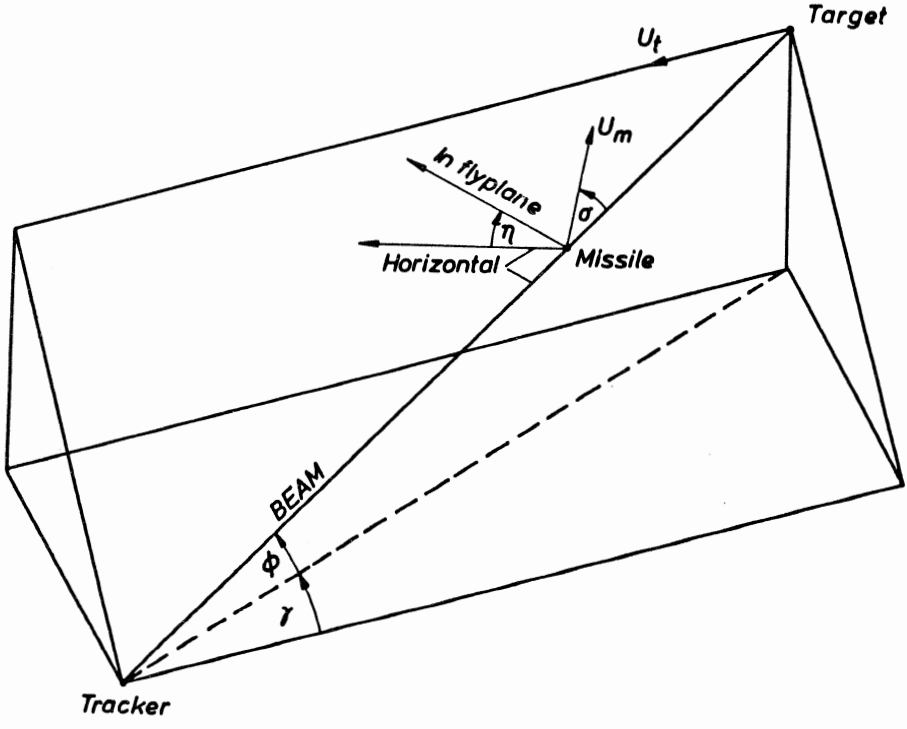


FIG 7.4-2 Flyplane geometry, a three dimensional view

$$\therefore \dot{\sigma}_m = \frac{\dot{\theta}\,\frac{d}{dt}\left(\frac{R_m}{U_m}\right) + \left(\frac{R_m}{U_m}\right)\ddot{\theta}}{\sqrt{1 - \left(\dot{\theta}R_m/U_m\right)^2}} \qquad (7.4\text{-}5)$$

Substituting this value of $\dot{\sigma}_m$ in equation 7.4-3 yields

$$f_1 = U_m\dot{\theta} + U_m \frac{\left\{\frac{d}{dt}\left(\frac{R_m}{U_m}\right)\dot{\theta} + \frac{R_m}{U_m}\ddot{\theta}\right\}}{\sqrt{1 - \left(\dot{\theta}R_m/U_m\right)^2}} \qquad (7.4\text{-}6)$$

We therefore need to know $\ddot{\theta}$ in addition to $\dot{\theta}$ and this can be obtained by differentiating $\dot{\theta}$. Since some noise will always be present a high frequency filter will also have to be used and this will mean in practice that the computed value of $\ddot{\theta}$ will always lag behind the true value. If missile latax or "bias latax" as it is sometimes called is computed in this way it is seen that the value of U_m and R_m/U_m as a function of time must be known with reasonable accuracy.

If now the target should manoeuvre perpendicular to the flyplane the missile will have to manoeuvre perpendicular to the flyplane in order to remain in the flyplane and hence in the LOS. The component of latax f_2 is obtained by taking the missile flight path rate about the beam axis and multiplying it by the component of missile velocity perpendicular to the beam $U_m \sin \sigma_m$. Hence

$$f_2 = U_m \sin \sigma_m \, (\dot{\gamma} \sin \phi + \dot{\eta}) \qquad (7.4\text{-}7)$$

where $\sin \eta = \dot{\phi}/\dot{\theta}$ and $\cos \eta = \dot{\gamma} \cos \phi/\dot{\theta}$ and either of these two relationships can be used to give η and hence $\dot{\eta}$.

The components f_1 and f_2 must now be resolved into "up-down" and "left-right" commands to the missile. Unfortunately our problems are not yet over as the missile's axis system will not coincide with the tracker's; there is an "orientation" problem and this is discussed separately. However, it is now clear that in order to compute feedforward commands we will need a digital computer. It should be noted that the computation of feedforward commands is complicated only for fast moving targets. Consider now a semi-automatic anti-tank system. Let us assume that the missile velocity is about 250 m/sec. It is unlikely that the target speed will exceed 10 m/sec. Missile latax will be very low and body to beam angles will be negligible. The reader can easily check by referring to equations 1.3-1, 1.3-2 and 7.1-1 that when the missile speed is much greater than that of the target bias latax is accurately given by

$$f_1 = 2U_m \dot{\theta} \qquad (7.4\text{-}8)$$

In fact, if the missile speed is roughly constant and it is assumed that target movement is largely restricted to the horizontal plane the feedforward command will be proportional to the azimuth tracking rate which can be measured by a tachogenerator mounted on the launcher tripod.

Before leaving the subject of feedforward terms it should be explained why in Figs 7.2-1, 7.2-3 and 7.2-4 it is inferred that feedforward is not employed in beam-riders. Now, one can compute on the ground the bias latax required in any one of these systems but in a beam rider the angular error signals are derived in the missile itself and there is no need for a command link from ground to missile. If a separate transmitter and receiver (in addition to the guidance receiver already in the missile) are used then feedforward can be employed. But such a complication is not usually resorted to; alternative methods would be to steer the riding beam ahead of the target by an amount proportional to the latax required. Conversely one could modulate the beam in some way so that the receiver in the missile knows that

some latax is required even if it is in the centre of the beam.

7.5 PHASING ERROR AND ORIENTATION DIFFICULTIES

We have seen in section 6.11 that if the missile is rolling there is mechanical cross-coupling between the two lateral autopilots and for a small servo phase lag ϕ_s at the roll frequency the effect on the autopilots is to a first approximation a loss of phase margin equal to ϕ_s. But this is not the end of the matter; each autopilot is a part of one of the guidance loops and if as a result of one or more of several causes part of the demand from one angle channel receiver is implemented in the other channel we say that a phasing error exists. Instead of the two guidance loops being entirely independent of each other they are now cross-coupled and Fig 6.11-2 represents the two guidance loops if the "inputs" and "outputs" are the target and missile directions respectively, Y_1 represents the guidance functions and the autopilot, Y_2 represents the kinematic closure of the loop and Y_3 is unity. To illustrate the effect consider a specific example where one of the demands is for a lateral acceleration of 150 m/sec^2 and there is a servo phase lag of 0.1 rad at the missile roll frequency assumed constant. Since $\sin \phi_s \simeq 0.1$ and $\cos \phi_s \simeq 1.0$ it follows that the gain of the guidance loops is not significantly affected and that approximately 15 m/sec^2 latax is developed in a direction perpendicular to the intended one i.e. in the other channel. If the guidance loop gain or stiffness is 10 say then the result of this is to cause the missile in the steady state to be 1.5m off the LOS in a direction perpendicular to the demand. Without the stiffness of the other channel to contend with the missile would merely accelerate out of the LOS at 15 m/sec^2. To a good first approximation the inaccuracy in guidance due to a demand for lateral acceleration f_d and a phasing error of ϕ radians is $f_d \sin \phi/K$ where K is the guidance loop stiffness. In addition, again to a first approximation, this phasing error ϕ will reduce the *guidance* loop phase margin by ϕ. It is important to note that the effect of phasing error could be really serious in systems where g demands are large e.g. surface-to-air or air-to-air systems. We now consider the other contributions to phasing error. We rely on the angle channel receiver up-down axis coinciding with the roll gyro vertical axis. Any difference here before launch will set up a phasing error. There may be a small uncaging error and the gyro will drift especially during boost. Now a really good resolver measuring to 10 minutes of arc or so is an expensive precision instrument and a small cheap

and relatively crude instrument will probably be used. If an induction resolver is used then some electronic circuitry will be required to convert the resolver outputs to d.c. signals. These three additional items will, between them, produce an error in amplitude and an error in angle. This angular error is another contribution to phasing error. It is possible for cross-coupling to occur in the receiver itself but usually any contributions to phasing error from this source are small compared with those just discussed. It will be noted that in general all these contributions are unknown and random in nature. We do not know in which sense the missile is rolling, nor is the roll rate known, the gyro drift is unpredictable and unless the resolver plus electronics are specially calibrated we cannot know this contribution either. Hence since these errors are unknown it follows that there is nothing we can do to allow for them.

We now come to another possible contribution to phasing error but since it is possible to allow for it in the guidance computer it is sometimes referred to as a phasing error due to missile orientation. It is due to the fact that the missile axes do not coincide with the tracker axes. Consider now a missile launched straight down a beam and executing small manoeuvres about this axis. A left-right signal requires a manoeuvre in a horizontal direction perpendicular to the beam and because the missile body coincides with the beam the latax should be developed perpendicular to the beam. Likewise an up-down demand does not necessarily require a vertical manoeuvre; it requires a lateral acceleration in the vertical plane and inclined back at an angle ϕ due to the elevation of the tracker. But since the missile ox axis is inclined to the horizontal by an angle ϕ this interpretation of "up-down" causes no difficulties. Now, except in the unique case of a target heading straight for the tracker the missile will be pulling latax in the flyplane and its velocity vector will be at an angle σ_m to the LOS as already discussed. This has been allowed for in deriving equation 7.4-6. However if enough computing facilities are available we should allow for the fact that latax is developed perpendicular to the body and not perpendicular to the velocity vector i.e. we need to compute the body incidence associated with a given latax; this implies that the missile speed and height must be known. In addition if the missile pitches *and* yaws after launch the gyro gimbal system will become distorted and as a result the roll angle will be incorrectly interpreted. This means that unless the correct axis transformations are done in the computer an additional phasing error will occur. We

have already noted the extent of this distortion in section 5.3. In making this axis transformation in the computer any toe-in from the launcher and superelevation of the missile should be allowed for. What we really want to know is the angle between the roll gyro inner gimbal and the flyplane.
It is important to note that it is phasing error due to roll rate and fin servo phase lag only that tends to unstabilise the autopilot, since this occurs inside the autopilot closed loop. Phasing error due to any cause whatever occurs inside the guidance loop and therefore the total phasing error affects the guidance loop.

7.6 THE EFFECT OF A DIGITAL COMPUTER INSIDE THE GUIDANCE LOOP

A relatively simple semi-automatic anti-tank system does not require a digital computer in the guidance loop; the "times range" term in the missile guidance loop can be engineered easily by analogue circuits. A high performance surface-to-air system designed for area rather than point defence will almost certainly employ a digital computer to work out the feedforward commands and since these have to be added to the compensated angle channel receiver signals the whole process of computing the demands for lateral acceleration is usually done digitally.
If the computer up-dates at intervals of τ seconds then an analogue to digital converter is required at one interface and a digital to analogue converter plus a hold or clamp circuit is required at the other. The effect on the transmitted information is illustrated in Fig 7.6-1. If a sinewave of frequency ω rad/sec is updated many times during a cycle the amplitude is not

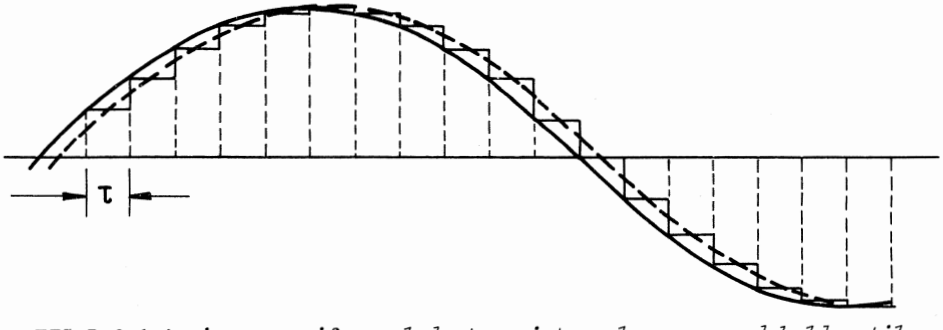


FIG 7.6-1 A sine wave if sampled at an interval τ secs and held until updated results in an approximate sinewave displaced in phase

sensibly affected but there is a shift in phase. Since the phase shift due to a time delay τ is $\omega\tau$ then it follows that the mean phase shift due to discontinuous information is $\omega\tau/2$. However, the computer will take some time

to do the necessary calculations and if this computing time is τ_c the total effective time delay due to the computer is $\tau_c + \tau/2 = \tau'$. The effect on the guidance loop is therefore a destabilising one. If the guidance loop open loop gain-crossover frequency is ω_c rad/sec then we must add an additional phase lag of $\omega_c\tau'$ at this frequency. For example if the guidance loop gain crossover frequency is 5 rad/sec, the data rate is 100/sec and the mean computing time is 5 millisecs then the phase lag at gain crossover due to the computer is 5 (.005 + .01/2) = 0.05 rad or nearly 3^0. Since such a loss of phase is barely acceptable in a guidance loop, it follows that fast computing times and high data rates are essential if computers are employed *inside* closed loop systems.

Consider now the method of implementing phase advance digitally, i.e. it is necessary to represent the function

$$\frac{u}{e} = \frac{1 + Ts}{1 + \alpha Ts} \tag{7.6-1}$$

in discrete form. This implies that the equation

$$u + \alpha T\dot{u} = e + T\dot{e} \tag{7.6-2}$$

must be implemented when e and u are available only in sampled form. Taking e_{n-1}, u_{n-1} and e_n, u_n to be the values of the continuous signals e, u at the $(n-1)$th and nth sampling instants then the simplest approximations to $\dot{e}$, $\dot{u}$ would be

$$\dot{e} \simeq \frac{e_n - e_{n-1}}{\tau}, \quad \dot{u} \simeq \frac{u_n - u_{n-1}}{\tau} \tag{7.6-3 and 7.6-4}$$

where τ is the sampling interval. In practice, more complicated expressions are often used although the simple approximations used here will serve for illustration. To obtain a value for u_n we use

$$u_n = u_{n-1} + \dot{u}_{n-1}\,\tau \tag{7.6-5}$$

and hence obtain the following recursive relation

$$u_n = c_1 e_n + c_2 e_{n-1} + c_3 u_{n-1} \tag{7.6-6}$$

where the coefficients c_1, c_2, c_3 are given by $1/\alpha$, $(\tau/T - 1)/\alpha$, $1 - \tau/\alpha T$ respectively.

In other words, the current output (u_n) of the digital network is dependent on the current value of the input (e_n) and on past values of both the input and the output (e_{n-1}, u_{n-1}). We can now consider the errors that may arise when implementing a continuous network in digital form. All these errors arise from the fact that, in practice, digital information can only be represented to a certain degree of accuracy which is determined by the number

of bits (word length) that is available.

This word length limitation will affect:

(1) The accuracy of the sampled signals at the input and output and results in the signals being confined to a range of discrete levels.

(2) The accuracy with which the network coefficients (e.g. c_1, c_2 and c_3 in equation 7.6-6) can be stored.

(3) The accuracy with which the arithmetic operations can be performed. For example, in implementing equation 7.6-6, it is necessary to multiply c_1 by e_n. If both these quantities are represented by a m-bit word then the product would need a 2m-bit word in order to sustain accuracy. To store this product as a m-bit word it would be necessary to dispense with the m least significant bits.

All the above inaccuracies, resulting from limited word length, can be considered as *quantisation* errors although the terms *truncation* or *round-off* are usually applied to the quantisation errors that occur when arithmetic operations are performed.

The effect of the quantisation errors on the sampled signals is similar to additive white noise and is undesirable for that reason alone. However, more serious consequences can occur at low signal levels where the quantisation effects tend to dominate the true nature of the signal and in these circumstances the behaviour of the network within the guidance loop can diverge considerably from its continuous counterpart.

The effect of inaccuracy in the network coefficients is dependent on the way in which the network equation is implemented or programmed. Similarly, the effect of truncation or round-off error is dependent on the mode of arithmetic used, i.e. whether fixed or floating point. Many of these effects are treated in detail in the literature and, in particular, reference 3 is recommended for its scope and relevance to the guided weapon field. However, one aspect of quantisation error will be considered here in greater detail in order to illustrate that transferring from continuous to discrete systems is not necessarily a straightforward procedure and that due account should be taken of limited word length effects during the earliest stages of system design.

Following an example treated in somewhat more detail in the above reference, we will assume that not all the coefficients of equation 7.6-6 can be represented exactly and, specifically, that the coefficient c_3 is stored as c_3' which is in error by an amount ε.

i.e. $$c_3' = c_3 + \varepsilon$$

Considering steady state conditions, where the input and output of the digital network remain at constant values such that $e_n = e_{n-1} = e_{n-2}$ ... etc and $u_n = u_{n-1} = u_{n-2}$ ... etc, then, from equation 7.6-6, we have

$$u_n = c_1 e_n + c_2 e_n + c_3' u_n$$

Thus $\frac{u_n}{e_n} = \frac{c_1 + c_2}{1 - c_3'}$ is the steady state gain of the network. Substituting for c_1, c_2 and c_3' gives

$$\frac{u_n}{e_n} = \frac{\frac{1}{\alpha} + \frac{\tau/T - 1}{\alpha}}{1 - (1 - \tau/\alpha T) - \varepsilon}$$

$$= (1 - \varepsilon\alpha T/\tau)^{-1}$$

$$= 1 + \varepsilon\alpha T/\tau + ...$$

It is seen that the steady state gain (which should be unity in this example) is dependent on both ε and τ and that far from improving the accuracy, a reduction in the sampling interval τ would exacerbate the situation. This result has arisen simply because of the limited word length available to represent c_3.

7.7 SOME NUMERICAL EXAMPLES ON THE ESTIMATION OF GUIDANCE ACCURACY

A medium range surface-to-air beam-riding system

If reference is made to Fig 7.2-1 it will be seen that two closed loop systems have to be designed, the target tracking loop and the missile guidance loop. Let us assume that the system is designed to hit approaching aircraft and missiles in the speed range M = 0.7 to M = 2.0 and that the minimum range is 6 km (largely because the initial dispersion of the missile is considerable due to the fact that it carries discarding boosts) and the maximum range is 40 km. We have already established in chapter 1 that a reasonable approach to designing the tracker servo loop is to include an electronic integrator and to reduce the lags due to the motor and load by designing a high gain rate feedback loop. We found that the minimum noise bandwidth was obtained when the lags in the rate loop were negligible and the closed loop damping ratio was 0.5. Let us determine the tracking error due to a constant speed straight flying target at minimum range assuming θ = 45, see Fig 1.3-1 and equation 1.3-3. The tracker closed loop transfer function well approximates to:

$$\frac{\theta_a}{\theta_t} = \frac{s/3 + 1}{s^2/9 + s/3 + 1}$$

Since $\ddot{\theta}_{max} = U_t^2/R_{min}^2$ and $\theta_e \simeq \ddot{\theta}/\omega_n^2$

$$\theta_e = \frac{(0.7 \times 340.3)^2}{6000^2} \times \frac{1}{9} \text{ for } M = 0.7 \text{ target}$$

$$= 0.000175 \text{ rad}$$

$$= 1.05 \text{ m at the target}$$

and since the sight line angular acceleration is proportional to (target speed)^2 this is 8.58 m at the target for the $M = 2$ target. We note however that since sight line acceleration $\alpha\ 1/R_t^2$ all other things being equal we would expect the tracking accuracy in metres at the target to improve with range. We now consider the effect of thermal noise. For a particular receiver we find that for a non-glinting 4 m^2 eachoing area target the spectral density of the thermal noise at 40 km range is given by

$$K_a^2 = 3.3 \times 10^{-9} \text{ rad}^2/\text{rad/sec}$$

We have already established in section 1.4 that the effective noise bandwidth of such a system is $\pi\omega_n$ rad = 9.42 rad. The mean square beam jitter is therefore given by

$$\sigma^2 = \frac{9.42 \times 3.3}{10^9} \text{ rad}^2$$

In metres at the target this is

$$\frac{9.42 \times 3.3}{10^9} \times 40^2 \times 10^6 \text{ m}^2$$

$$= 49.7 \text{ m}^2 \text{ or } 7.05 \text{ m r.m.s.}$$

This is for a 4 m^2 echoing area target but would be considerably greater for a small target such as a missile with an effective echoing area of 0.1 m^2. On the other hand the effect of noise will improve with a reduction in range. At 10 km range the signal to noise power should have improved by a factor of 4^4 reducing the r.m.s. beam jitter by a factor of 4^2; in metres this is an improvement by a factor of 4^3.

If the target is glinting the r.m.s. beam jitter due to glint is likely to be close to the r.m.s. value of the glint itself unless the servo bandwidth is very low indeed or frequency agility is used, see Table 1.6-1. Summarising we are saying that accurate tracking is a difficult problem with fast targets associated with short ranges due to the high angular accelerations, and as the range increases thermal noise tends to take over as the major source of

inaccuracy. Tracking inaccuracy due to glint is not a function of range. We now consider the design of the guidance loop. Since the missile is designed for medium range it is likely to be heavy, 1000 kg or more. For geometrically similar missiles, inertias $\alpha\ \ell^5$ and control surface moments $\alpha\ \ell^3$. Hence large missiles will tend to have lower weathercock frequencies. Suppose it is feasible to design a lateral autopilot whose dynamics approximate to that of a quadratic lag defined by ω_n = 12 and μ = 0.6. The maximum phase lag permissible at guidance loop (open loop) gain cross-over frequency has been discussed in section 6.2. If we restrict this to 15^0 as suggested it means that the maximum d.c. gain we can use is about 3; the open loop gain cross-over frequency is always greater than the square root of the gain when there are two integrations, due to the effect of the phase advance networks on gain. Let us therefore consider the sort of miss that would be involved if the guidance open loop gain is set to 3. One typical engagement will suffice to highlight the problem. Assume U_t = 250 m/sec and that the target is flying straight such that the crossing distance is 6 km. If the missile is gathered at a slant range of 5 km at which time the target slant range is 13.5 km and the missile speed is constant at 500 m/sec, some geometric manipulation reveals that interception should occur at a slant range of 10.86 km, 12 secs after gathering. The reader may wish to verify that the missile latax at impact is 18.2 m/sec^2. Since the loop gain is 3 it means that the miss due to "bias latax" is approximately 18.2/3 $\simeq$ 6m. Accurate simulation shows that this assumption results in a slight overestimate of the miss if the latax is increasing. We have noted this effect when estimating the error of tracking servos in Figs 1.3-2 and 1.3-3. If interception is at a shorter range than this or the target speed is greater the missile latax will have to be greater, resulting in a larger miss. Can we structure the loop in a different way to reduce this bias miss? Consider now the use of proportional plus integral control rather than proportional control. This will result in a loss of phase margin so let us reduce the d.c. gain to unity say. If an integrating time constant of 2 is adopted it means that if the missile error off the beam is constant at x metres the autopilot output is increasing at a rate of $x/2$ m/sec^3. The rate of increase of missile latax during the last few seconds of the engagement is about 0.75 m/sec^3. This is easily determined by calculating the necessary latax 1 second before impact and comparing it with the latax required at impact. The miss should now be 0.75 x 2 = 1.5m

hence we have reduced the bias latax miss by a factor of 4 to 1. In addition we have reduced the guidance d.c. gain by a factor of 3 and have therefore reduced the effective noise bandwidth; the r.m.s. noise into the missile servos is also decreased considerably. If a double phase advance network is used to stabilise the loop the complete guidance loop in transfer function form is shown in Fig 7.7-1.

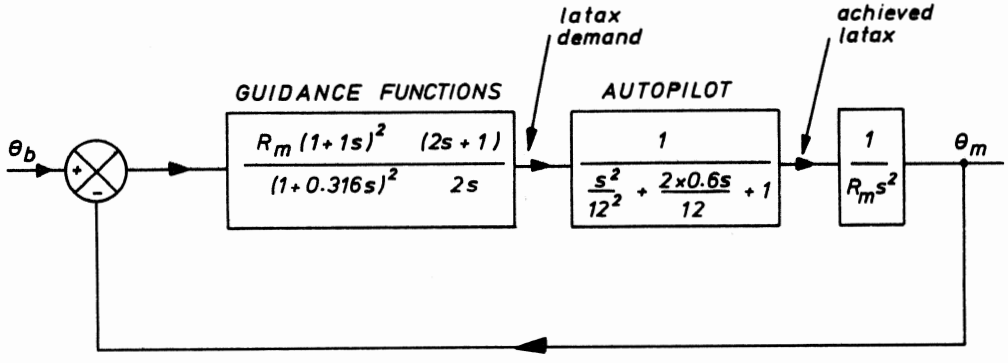


FIG 7.7-1 A beam-riding guidance loop. Block diagram for stability analysis

The kinematic closure is shown to be two integrations and a $1/R_m$ term as we have already established in section 7.3 that this is a reasonable assumption for stability calculations. The Bode diagram for this guidance loop is given in Fig 7.7-2 and this reveals a phase margin of 35^0 and a gain margin of 10.5 dbs.

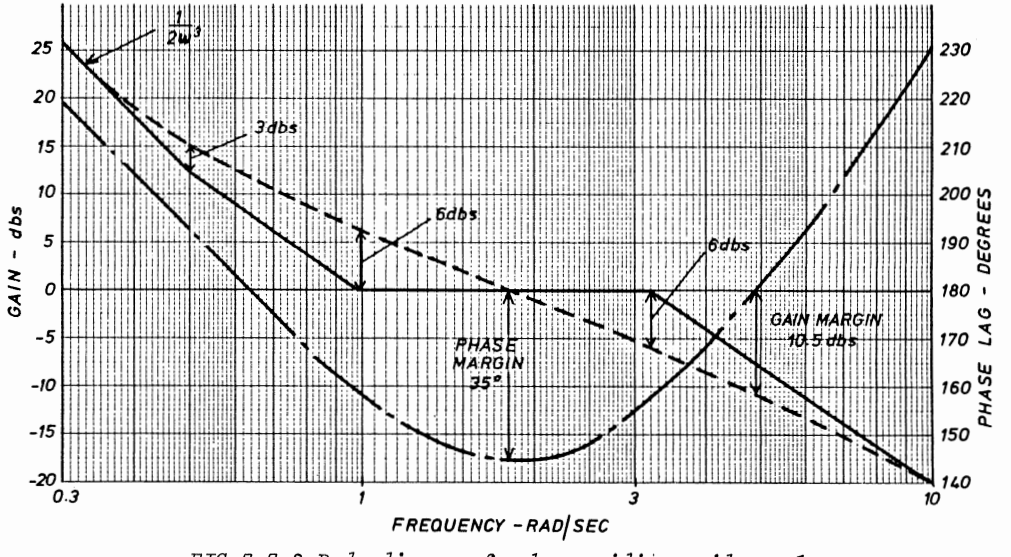


FIG 7.7-2 Bode diagram for beam-riding guidance loop

The actual miss of the missile due to thermal noise can be evaluated by multiplying the tracker transfer function by the guidance loop transfer function and calculating the equivalent noise bandwidth to thermal noise of the combination. Integrals for transfer functions up to tenth order are given by Newton Gould and Kaiser (4). We go through the same procedure for glint except that we include an additional first order filter to account for the fact that glint is not "white" noise but is "coloured" noise. Detailed analysis shows that the r.m.s. miss due to noise is not much less than the beam jitter. This is due to the fact that the *mean square* output of a linear filter due to white noise is proportional to the noise bandwidth and therefore the r.m.s. output is proportional to the square root of the noise bandwidth. Since the damping in this guidance loop is on the low side the equivalent noise bandwidth is increased. An important feature of a guidance loop with an extra integrator is that any bias in the autopilot due to fin bias, or instrument biases results in zero miss, provided of course that the biases remain constant or change only slowly.

The above example is a good illustration of the compromise that has to be made in all guidance loop designs. We require high gain for steady state accuracy but low gain for stability and low miss due to noise, and low gain to avoid saturation of the missile servos. In this particular design because engagement times are typically in the range 15-80 seconds the addition of a slow integrator in the loop has increased the steady state accuracy considerably; nevertheless some 16^{o} of phase margin has been sacrificed by using proportional plus integral control leaving us with a rather underdamped guidance loop.

A short range semi-automatic surface-to-air CLOS system

In this case we assume the human operator tracks the target with the help of a servo-controlled optical system. The question we now ask is, how accurately can a trained operator track an aircraft? It is generally agreed that a velocity servo gives him most assistance; a given controller position will result in a constant angular velocity of the sight. But an approaching target results in an accelerating sight line. If therefore the system is specifically designed for approaching targets it will help the operator to add an integrator with a low gain so that a constant input will produce a slowly accelerating output. In any particular engagement the integrator gain will be either too high or too low; an extensive programme of realistic simulations will be necessary to arrive at the best compromise. Also some

phase advance will tend to offset the missile lag. Several systems have been designed which claim a tracking accuracy of about 0.2 to 0.3 milliradians r.m.s. in the flyplane and better than this perpendicular to the flyplane at ranges of 2-6 km, but much will depend on how near the target speed and crossing range are matched to the controller time constants.

If the target is tracked with the aid of a vidicon and a flare on the missile a collimation problem will exist. Collimation errors will arise not only because of differential expansions due to temperature changes but also due to a drift in the electrical null axis of the target tracker. Collimation must be capable of being checked in the field.

In a short range system missile g requirements can be very high; many systems are designed for a maximum of 25g or even greater. If the impact range is from 1 km to 6 km and the engagement time is anything from 2 to 12 seconds it is clear that missile g requirements can be changing very rapidly during flight. To keep the bias latax miss to say 2m if the missile latax is 200 m/sec^2 requires a guidance loop gain of 100 or alternatively a lower d.c. gain plus a fast integrator. The first option would pose severe stability problems and would almost certainly result in the fin servos being saturated with noise. The second option would probably be equally impracticable due to the unstabilising nature of a fast integrator. The alternative, we have seen, is to use feedforward terms. Any bias latax miss is now due to any imperfections in computing and implementing these feedforward commands. In other words, if we adopt a guidance loop gain of say 15 an error of x m/sec^2 in computing and implementing this command will result in a miss in the flyplane of $x/15$ m. If the achieved latax due to feedforward is greater than the correct value it will fly ahead of the LOS and vice-versa. Put in another way if the feedforward command results in 90% or 110% of the correct command being implemented the guidance loop provides the other +10% or -10% respectively. In either case the guidance accuracy on this account is increased by a factor of 10 to 1. Note that the use of feedforward commands makes no difference to the miss due to phasing error.

Any biases in the autopilot will cause a miss unless there is an electronic integrator in the loop. It is customary to estimate the total likely bias in the autopilot in m/sec^2 and this will usually be the same figure for each plane for a symmetrical cruciform missile. This bias divided by the guidance loop gain will be the miss due to this cause.

And finally, there will be the miss due to thermal noise in the guidance loop

which is usually small at such short ranges. If now the individual contributions to guidance inaccuracy are uncorrelated and can be assumed to have a normal distribution we can estimate the total mean square miss by obtaining a mean square sum. This assumption is perfectly valid if the input is white noise; provided the system can be regarded as a linear filter the output will have a normal distribution with zero mean value. The assumption is also valid if the input is glint and it is assumed that its spectrum is white noise passed through a linear filter. We are on less certain ground when considering biases in say the autopilot. However, an assumption often made is that the bias from a large batch of accelerometers, rate gyros etc has zero mean value and the distribution of individual biases is a normal one with a 1σ value equal to half the manufacturing tolerance.

Before tabulating the various contributions to miss distance we should perhaps consider some autopilot imperfections in some detail. Assume that the lateral autopilot parameters are those used in the example given in section 6.3 i.e.

autopilot nominal gain : 1m sec^{-2} $volt^{-1}$
accelerometer nominal gain: 0.8 volts m^{-1} sec^2
rate gyro nominal gain : 30 volts rad^{-1} sec

If the accelerometer full scale deflection is 30g and the tolerance is ± 2% of full scale or 5% of the actual value whichever is the greater then a zero demand would give a 1σ bias of ± .01 x 30 x 9.81 x 0.8 = 2.35 volts and this in turn produces an autopilot bias of 2.35 m/sec^2. A demand for 25g would give a 1σ bias of 0.025 x 25 x 9.81 x 0.8 ≡ 4.9m/sec^2. The rate gyro full scale reading is 5 rad/sec and the tolerance on bias is ± 2% of full scale, and *in addition* there is a gain tolerance ± 10% of the actual value (largely because it is blast started and the gain varies as the gyro wheel slows down). Hence when the missile is not turning (0g latax) the bias is 3 m/sec^2 and when the missile is pulling 25g there is a rate of turn of 25 x 9.81/500 = 0.49 rad/sec if the velocity is 500 m/sec. Hence there is likely to be an additional bias of 5 x 30 x 0.05 x 0.49 ≡ 3.7 m/sec^2. A $\frac{1}{2}^o$ fin bias (1σ value) *open loop* would result in a bias of 11.4 m/sec^2 at the nominal velocity of 500 m/sec but with instrument feedback this is reduced to 1.3 m/sec^2. Table 7.7-1 therefore summarises the position for zero and maximum latax being pulled in the flyplane. The 1σ latax error in m/sec^2 has been divided by the guidance loop gain of 15 to give the 1σ miss in metres. It has been assumed that impact occurs at 4 km. A somewhat arbitrary contribution to miss distance due to tracking the target has been shown as this will depend on many factors

TABLE 7.7-1 1 MISS DISTANCE CONTRIBUTIONS DUE TO FAULTY COMPUTATION AND EXECUTION OF FEEDFORWARD COMMANDS, TRACKING ERRORS, PHASING ERROR ETC

SYSTEM COMPONENT	CAUSE OF ERROR	2σ ERROR MAGNITUDE OR TOLERANCE	1σ miss in flyplane-metres		1σ miss perpendicular to flyplane-metres	
			Og	25g	Og	25g
Ground Computer	Tachnogenerator inputs uncertainty in missile range and velocity, word length	7% of actual value or 2% of full scale whichever is the greater	0.16	0.57	0.16	
Missile Receiver	Time comparator associated with decoder	2% of actual value or 1% of full scale	0.08	0.16	0.08	0.08
Resolver and Resolver Circuitry	Manufacturing and component tolerances	± 5% of actual value	–	0.40	–	–
Autopilot gain and biases	Variation in gain due to changes in height and speed	± 10% of the nominal gain	–	0.82	–	–
	Accelerometer bias and gain tolerance	± 5% of the actual value or ± 2% of full scale	0.13	0.33	.13	.13
	Rate gyro bias and gain tolerance	± 10% of the actual value. 2% of full scale	– 0.10	0.24 0.10	– 0.10	– 0.10
	Fin servo bias	± 1°	0.09	0.09	0.09	0.09
Phasing error	Gyro mounting uncaging and drift	± 5°				
	Resolver	± 3°				
	Resolver circuitry	± 2°				
	Missile roll rate 2σ value 10 rad/sec	3.2°				
	Total 2σ phasing error	7.0°	–	0.02	–	1.00
Human Operator	Target tracking inaccuracy - bias plus noise	Varies with range, sightline rate, sightline acceleration	0.60	1.20	0.30	0.50
Tracking head assembly	Collimation error	2 milliradians	0.40	0.40	0.40	0.40
Guidance receiver	Thermal noise due to tracking missile	Will vary with range	0.20	0.20	0.20	0.20
	R.m.s. miss – metres		0.78	1.73	0.60	1.23

as already indicated. However if the missile is pulling high latax it means that the sight line rates and accelerations are high; target tracking tends to be less accurate under these conditions.

It is seen that in this hypothetical system a major contribution to miss distance is the inaccuracy in target tracking. At high g's there is a large contribution due to phasing error in the plane perpendicular to the flyplane; also under these conditions there is a significant contribution due to the variation in the autopilot d.c. gain. One way of improving this system aspect would be to vary the autopilot forward path gain as a function of missile speed and height. In any case system accuracy is bound to degrade somewhat if the missile is pulling high g's.

A system using a differential tracker

The contributions to miss distance due to faulty feedforward apply equally to systems using differential trackers. However, errors due to tracking the target are largely eliminated but there can be a small miss due to the two gains not being precisely equal, see Fig 7.2-4. Collimation error does not arise, but there will now be a contribution due to glint.

Some general conclusions on guidance accuracy

There are many techniques available to track targets other than the use of radar. The contributions to miss distance will clearly vary from system to system. For instance if the infra-red emission of the target is sensed by a suitable detector, the phenomenom of glint will not arise. Finally, it should be stressed that although good estimates of guidance accuracy may be obtained in this way, it is usual to check these results by a detailed simulation. If the simulation is a two-dimensional one the aerodynamics are reasonably linear and no serious saturation occurs in any of the elements, the results are likely to be very similar to the estimates obtained using the methods above. However, if the fin servos become saturated by noise near interception, or if the aerodynamics are very non-linear, or if there is much aerodynamic cross-coupling and a complete three dimensional simulation is carried out, the results may be quite different; the miss is usually considerably greater. However, designers are well aware of these dangers and seek to avoid them whenever possible.

REFERENCES

1. GRIDLEY D.O.A. An analogue computer study of the interception kinematic properties of boost-coast line of sight missiles. Royal Aircraft Establishment

Technical Report 66274 September 1966.

2. FLIGHT International 22nd April 1971.

3. FOX J.E. An introduction to the effect of word length on digital networks. British Aircraft Corporation report no ST 7135 May 1972.

4. NEWTON C.J., GOULD L.A., and KAISER J.F. Analytical design of linear feedback controls. John Wiley and Sons Inc 1957.

CHAPTER 8

HOMING HEADS AND SOME ASSOCIATED STABILITY PROBLEMS

8.1 INTRODUCTION

In the next chapter we consider the behaviour of systems using homing guidance and it is pointed out that the great majority of these systems use a particular kind of guidance law called "proportional navigation" which calls for a rate of change of flight path k times the rate of change of sight line where k is called the navigation constant. Before this can be implemented we have to compute the sight line rate and this is done by including a homing head in the missile. A homing head is merely a target tracker; when mounted in an airborne missile it is sometimes called a "homing eye" or a "seeker". In homing systems the sight line is defined as the direction of the line joining *missile* and target with respect to a space datum. Since we will be concerned only with differences or with rates, any space datum will suffice. For convenience Fig 8.1-1 is drawn to define the angles in one plane only but in

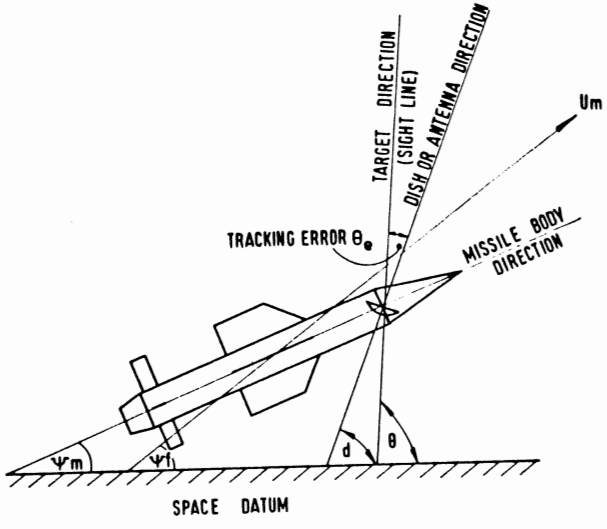


FIG 8.1-1 Definition of angles

practice we will be concerned with two components of sight line rates. One simplification which arises when the target tracker is mounted in the missile is that no roll resolution is required since the missile control axes (pitch

and yaw) can be made to coincide with the tracker up-down and left-right axes. Indeed if the missile should roll the guidance and control axes roll together. There will be no loss of generality if we confine our arguments to motion in one plane.

8.2 HOMING HEAD REQUIREMENTS

The accuracy of all target trackers is affected to some extent by the motion of the mounting. Since the homing head is mounted on the missile framework, it follows that there could be a tendency for "base" motion to contaminate the computation of sight line rate. It appears that some sort of gyroscope in the head is required to give an inertial reference. In all systems the homing head has to remain stabilised in space during a period when there is no guidance signal e.g. in a system when the target is tracked and illuminated by a special tracker (usually on the ground) and the homing eye detects some of the reflected energy. During the boost period the excessive motor efflux prevents the rear reference signal getting through and no guidance signals are possible.

Since the homing eye is mounted in a set of gimbals in the missile there is going to be limited angular freedom with respect to the missile. It is shown in the next chapter that in a well designed homing system the missile automatically develops a lead angle and, unless the target starts manoeuvring, the missile tends to fly straight near interception. Fig 8.2-1 shows the

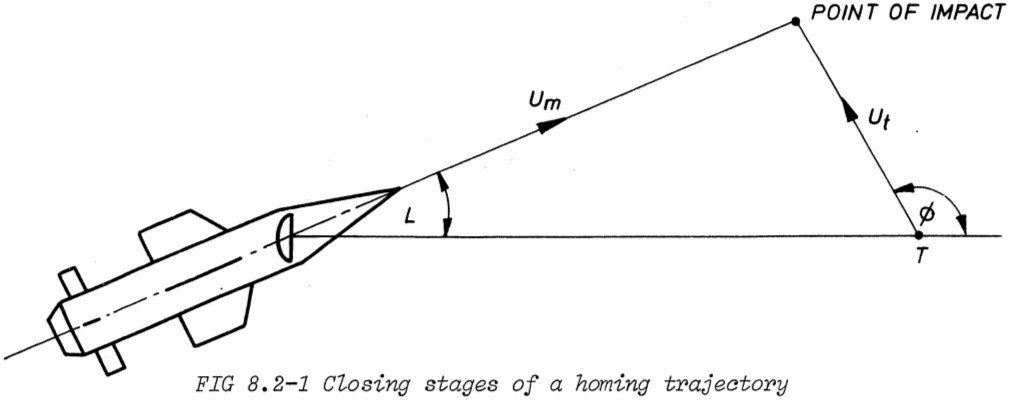


FIG 8.2-1 Closing stages of a homing trajectory

closing stages of an engagement. The homing head look-out angle for zero body incidence is the difference between the flight path direction and the sight line; the tracking error is normally a fraction of a degree. This look-out angle (L) for constant missile and target speeds U_m and U_t respectively can be deduced from

$$U_m \sin L = U_t \sin \phi$$

where ϕ is the angle the target velocity vector makes with the line joining the missile and target i.e.

$$\sin L = \sin \phi / n$$

where $n = U_m/U_t$. For a given speed ratio the look-out angle is a maximum if $\phi = 90^0$. If a system is designed for near head-on engagements or tail chases only ϕ cannot approach 90^0 in practice, but on the assumption that any geometry of engagement has to be accepted and we allow $\pm 15^0$ say for body incidence (i.e. $\psi_m \neq \psi_f$)the minimum angle that must be available mechanically in the gimbal arrangement is given in Table 8.2-1.

TABLE 8.2-1 MINIMUM GIMBAL ANGLE NECESSARY FOR A GIVEN SPEED RATIO

n	ϕ^0	L^0	Minimum gimbal angle
1	150	30	45
	135	45	60
	120	60	75) Not possible in
	90	90	105) practice
$\sqrt{2}$	150	18	33
	135	30	45
	120	38	53
	90	45	60
2	150	15	30
	135	21	36
	120	26	41
	90	30	45

It has been found that it is not practicable to design homing heads with gimbal angles more than $\pm 60^0$. Many designs are limited to about $\pm 40^0$. It is seen therefore that if broadside engagements are to be possible a speed advantage of at least $\sqrt{2}$ is essential.

And finally the homing head servos which keep the antenna pointing at the target must have a sufficiently high bandwidth. It is argued in the next chapter that a homing head bandwidth of between 1 and 2 Hz should be adequate for most purposes. However if there is likely to be a large and sudden disturbance due to the missile rolling and there is much mechanical coupling between the head and the missile body a high bandwidth servo is essential to keep the tracking accuracy well within the antenna beamwidth. With fast crossing targets the sight line rate may be considerable and the

servos must be able to move or "precess" the head at least at this rate.

8.3 SOME ELECTRO-MECHANICAL ARRANGEMENTS

A target tracker will keep itself aligned with its boresight pointing at the centre of reflecting area of the target. With a stationary mounting the accuracy will be limited by noise, biases and friction. If there is some base motion which tends to disturb the antenna a pointing error will result and the servos will drive in such a sense as to bring the dish back to its original position. In the absence of any base motion the supply voltages to the motors could be used as a measure of the antenna steady state angular rates, and for perfect servos the antenna rates will always be equal to the sight line rates. In practice of course there will be a dynamic lag in the antenna following the sight line. In the next chapter desirable homing head servo bandwidths will be discussed. Missile motion in pitch and yaw however, poses a problem. Depending on the degree of mechanical coupling between the drive motor and the head itself, missile motion will tend to disturb the head. The results of the servo restoring efforts will now be interpreted as due to target motion and spurious steering signals will be generated. Clearly a system is required which aims at 100% dish-body decoupling.

The electro-mechanical arrangement of a homing head designed at the Royal Aircraft Establishment is shown in Fig 8.3-1. The detector is space-stabilised by a gyro wheel which is spun up by releasing high pressure air from a small bottle. Precessing torques in the pitch and yaw planes are provided by two pairs of servo-operated solenoids. The c.g. of the detector and gyro assembly lies at the intersection of the gimbal axes. Such a system is very well decoupled from the body. There will be stray torques from the leads to the detector and some frictional forces from the bearings but these will be very small compared with the gyroscopic rigidity $J\omega$. An objection to the use of solenoids is that forces and resulting torques, and hence precessional rates are proportional to current squared and are a non-linear function of the air gap. To produce a steering signal proportional to sight line rate would need some non-linear compensating circuits.

A more conventional homing head designed by Marconi Space and Defence Systems Ltd is shown in Fig 8.3-2. Conventional permanent magnet d.c. motors and a total gear reduction of 200 to 1 are used to drive the head, but this arrangement leads to some dish-to-body coupling. Imagine an angular acceleration of the body. Since the motor is geared to the body it follows that if the head is to remain stationary the motor must accelerate. Also, if there is any

PRINCIPAL DATA	
WHEEL:-	
MASS.	0·7 Kg.
INERTIA.	1020 Kg. mm^2
ANGULAR MOMENTUM.	1·07 Kg m^2/S
SPEED.	10.000 r.p.m.
RUN UP TIME.	2 SEC.
PRECESSION RATE.	10°/SEC.
LOCK ANGLE.	± 40°
DURATION.	> 30 SEC.
ELECTRICAL POWER:	80 WATT. (MAX.)
SYSTEM MASS:	3·8 Kg. (APPROX.)
SERVO DEVELOPS MAXIMUM RATE FOR 1° MASALIGNMENT	
SERVO PRODUCES 50° PHASE LAG AND RESPONSE IS 3 DB DOWN FOR A SINUSOIDAL INPUT OF ± ½° AT 2 HZ.	

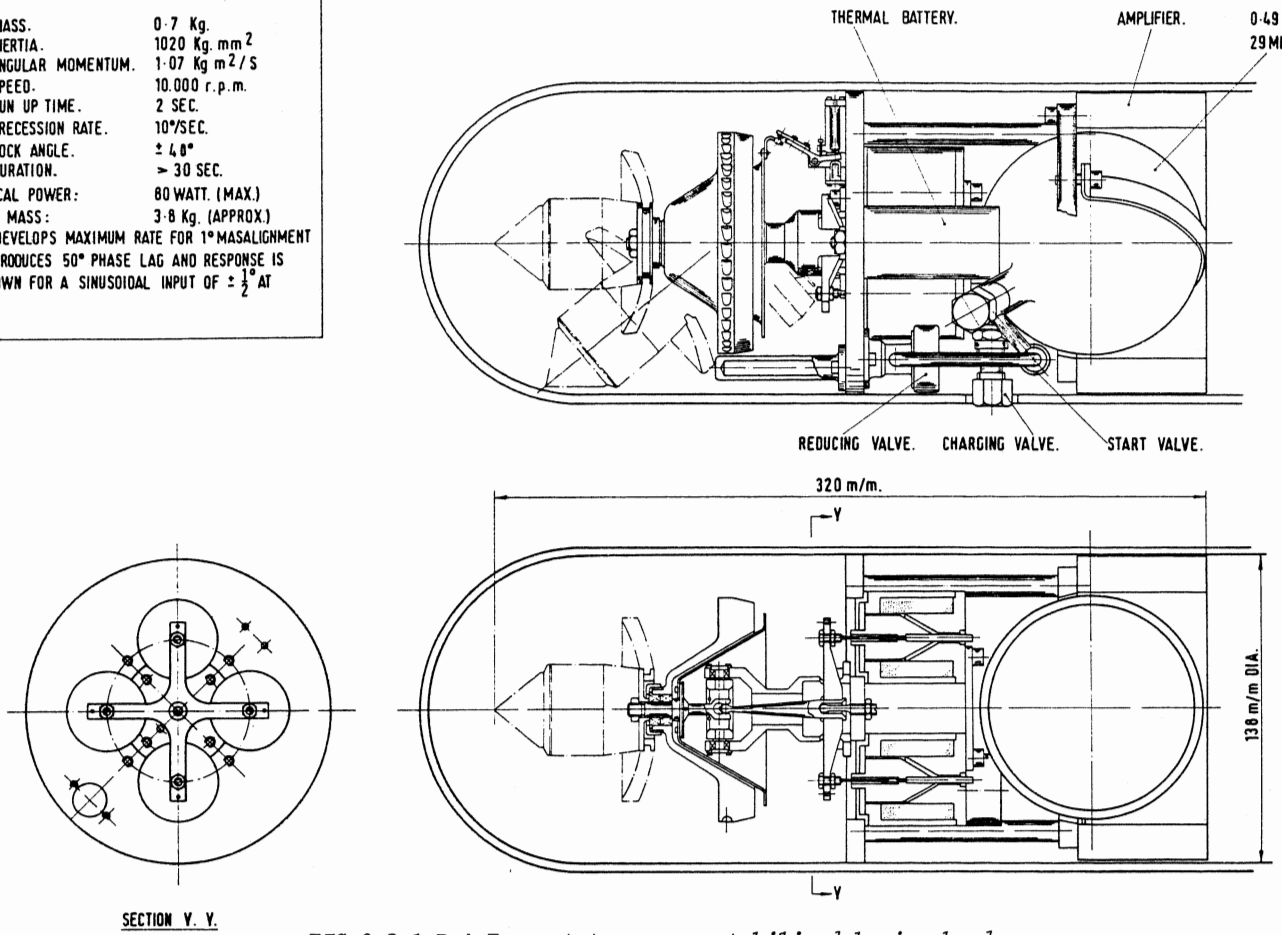
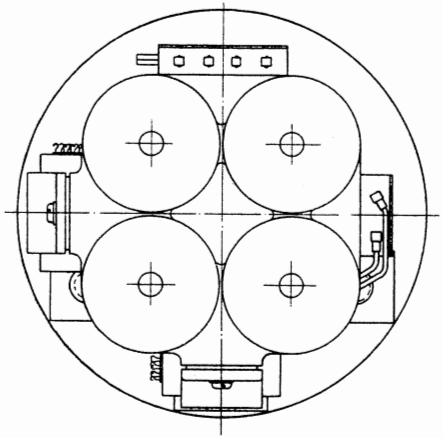
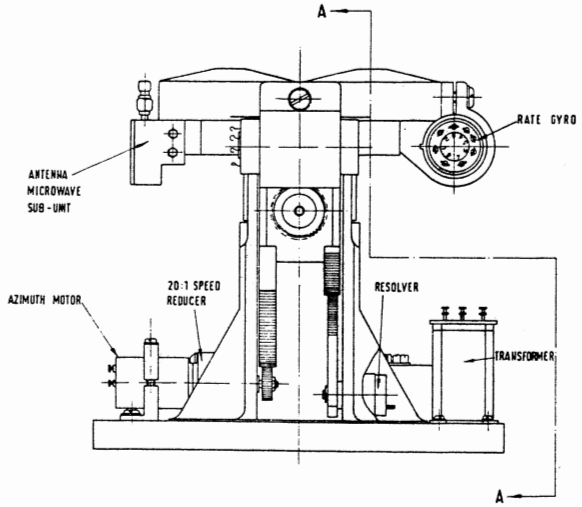


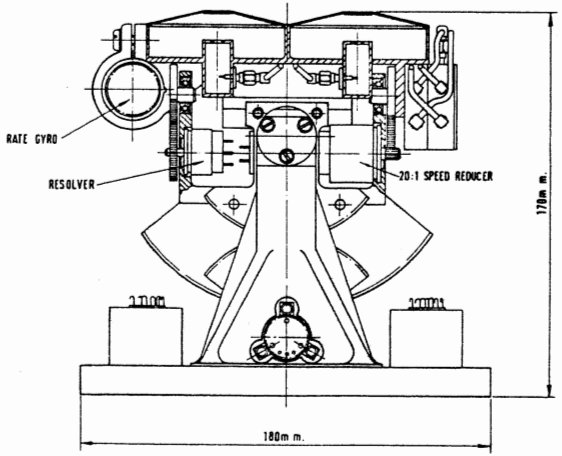
FIG 8.3-1 R.A.E. prototype gyro-stabilised homing head

MAXIMUM PRECESSING RATE 40°/SEC.
DC PERMANENT MAGNET SIZE 11 SERVO MOTORS.
MINIATURE RATE GYROS FROM NORTHROP OR ELLIOTT TYPE (GRH 4).
APPROX. WEIGHT OF PARTS SHOWN - 4·5 kg.
SERVO ERROR FOR MAXIMUM TRACKING RATE: 0·2°.
NOMINAL SERVO BANDWIOTH: > 10 HZ.





VIEW WITH GYRO CHOKE REMOVED



SECTION A-A WITH TRANSFORMER REMOVED

FIG 8.3-2 Marconi Space and Defence Systems Ltd homing head using d.c. motors and rate gyros

relative velocity between the motor rotor and the body a voltage will be generated in the motor windings. The design principle is akin to that shown in Fig 1.2-1 in that one relies on high gain inner rate loops using rate gyros largely to decouple the head from body motion. In some large homing heads hydraulic jacks are used to move the head and the mechanical coupling problem can become more severe. The additional techniques of "missile compensation" and "line of sight isolation" are discussed separately. It can be argued that the rate gyro outputs are not the best signals to represent sight line rates as in addition they contain the stabilisation signals which we do not want to know about. Ingenious schemes have been devised to mix the error signal with the rate gyro signal such that the resultant is the best representation of the sight line rate.

It will be noted that the sight line rates are those as seen by the homing eye and except when one looks straight ahead the homing eye axes will not coincide with the missile axes. If the outer gimbal is mounted with its axis of rotation in the sense oz and has rotated through α, and the inner gimbal through β from the straight ahead position, the resolutions which are necessary are

$$\dot{\theta}_{mz} = \dot{\theta}_{hz} \cos \beta$$

and

$$\dot{\theta}_{my} = \dot{\theta}_{hy} \cos \alpha + \dot{\theta}_{hz} \sin \alpha \sin \beta$$

where suffices m and h refer to missile and head respectively. This requires two sine-cosine resolvers and these are shown in position in Fig 8.3-2. There is a further complication in the gyro wheel stabilised version since a steady precessional rate about the outer gimbal axis is reduced by the cosine of angle of rotation of the inner gimbal i.e. by cos α; the gyro spin vector $J\omega$ is no longer orthogonal to the torque axis. To obtain an accurate measure of the precessional rate one has to attenuate the signal proportional to torqueing current by cos α. Nevertheless some systems which expect to have a large speed advantage do not do this resolution; the angles α and β are expected to be small and detailed simulation shows that the accuracy of homing is not significantly affected.

And finally, how does a homing eye acquire a target at all? There are two distinct methods. If there is a surveillance radar which has acquired a target as a result of an angular search and has stored its angular position the homing head can be pointed in the same direction by a simple 1 to 1 position servo link; this means that the surveillance radar and missile launcher have to be aligned. The alternative method is to make the head search for a target. In the case of the gyro wheel-stabilised version, if the

precessing motors are fed with currents 90^0 out of phase the head will describe an ellipse. If the incoming energy level reaches a certain level the precessing currents are inhibited; the homing head should then "lock on".

8.4 THE EFFECT OF RADOME ABERRATION

Consider a homing missile with the target tracker inside a protective cover called a radome, the external shape of which is determined mainly by aerodynamic considerations. The radome causes an angular displacement of the tracking aerial polar diagram and the angular difference between the apparent sight-line and the actual sight-line is called radome aberration. For a given radome this aberration is a function of the angle of look $\theta - \psi_m$, see Figs 8.4-1 and 8.4-2. If the aberration is not constant any missile motion will change the angle of look and apparent sight-line rates will be generated, which in turn will affect missile attitude. Small diameter missiles are at a disadvantage here since aberration is a function of wavelength-to-diameter. We have to ask

(a) Can this apparent dish-body coupling unstabilise the guidance loop?

(b) Can it degrade the system performance?

(c) How can one specify what sort of radome aberration can be tolerated?

For any value of $\theta - \psi_m$ the aberration angle can be defined by

$$\theta_{ab} = \sigma + \gamma\ (\theta - \psi_m) \tag{8.4-1}$$

where σ is not necessarily a constant and γ is the aberration slope at that point. Hence the angle θ' as seen by the homing eye is given by

$$\begin{aligned} \theta' = \theta + \theta_{ab} &= \theta + \sigma + \gamma\ (\theta - \psi_m) \\ &= \theta(1 + \gamma) + \sigma - \gamma\ \psi_m \end{aligned} \tag{8.4-2}$$

If the homing head produces a signal proportional to sight line rate and the autopilot a rate of change of flight path as a result of a command and both these subsystems are adequately modelled as quadratic lags then the block diagram is as shown in Fig 8.4-3. The effective forward d.c. gain has been combined as one number, k the navigation constant. This number for satisfactory interception of targets is usually in the range 2-6. Since the quantity σ is an effective input or disturbance to the system rates will be generated if it is not a constant. This will clearly affect system accuracy; a realistic simulation with a real homing head and real source incorporated in the system simulation is the only satisfactory way of assessing its affect. However, the aberration slope factor γ now provides a new feedback path and its effect on guidance loop stability will now be considered. The transfer function $\dot{\psi}/\theta$

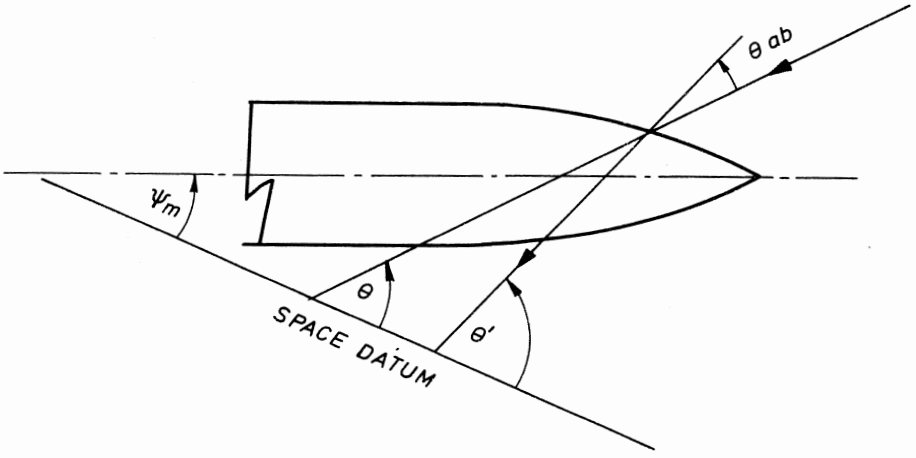


FIG 8.4-1 Radome aberration

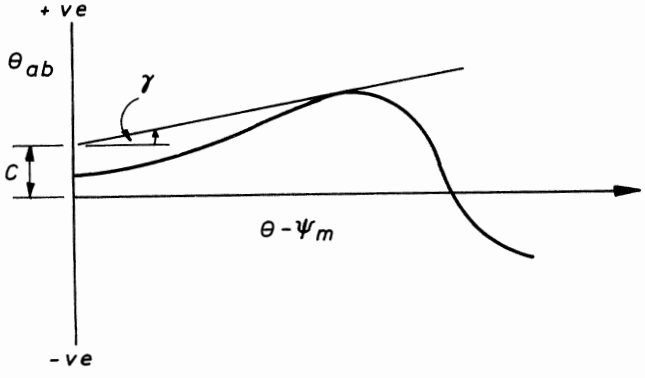


FIG 8.4-2 Aberration angle as a function of sightline-to-missile angle

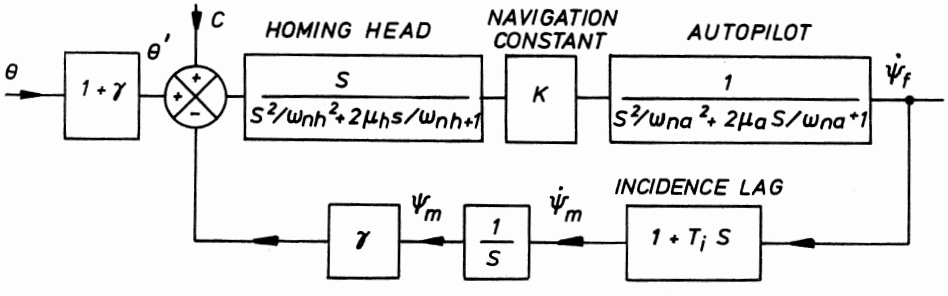


FIG 8.4-3 Parasitic feedback path caused by radome aberration

is now

$$\frac{\dot{\psi}_f}{\dot{\theta}'} = \frac{\dot{\psi}_f}{\theta(1+\psi)} \cdot \frac{1}{\dfrac{\left(\dfrac{s^2}{\omega_{nh}^2} + \dfrac{2\mu_h s}{\omega_{nh}} + 1\right)\left(\dfrac{s^2}{\omega_{na}^2} + \dfrac{2\mu_a s}{\omega_{na}} + 1\right)}{sk} + \dfrac{(1 + T_i s)\gamma}{s}}$$

which simplifies to

$$\frac{\dot{\psi}_f}{\dot{\theta}} = \frac{k\,\dfrac{1+\gamma}{1+k\gamma}}{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + 1} \tag{8.4-3}$$

where

$$a_4 = \left(\frac{1}{\omega_{nh}^2\,\omega_{na}^2}\right)/(1 + k\gamma)$$

$$a_3 = \left(\frac{2\mu_h}{\omega_{nh}\,\omega_{na}^2} + \frac{2\mu_a}{\omega_{na}\,\omega_{nh}^2}\right)/(1 + k\gamma)$$

$$a_2 = \left(\frac{1}{\omega_{nh}^2} + \frac{1}{\omega_{na}^2} + \frac{4\mu_h\,\mu_a}{\omega_{nh}\,\omega_{na}}\right)/(1 + k\gamma)$$

$$a_1 = \left(\frac{2\mu_h}{\omega_{nh}} + \frac{2\mu_a}{\omega_{na}} + k\,T_i\,\gamma\right)/(1 + k\gamma)$$

The simplest possible situation arises when the autopilot is very much faster than the homing head and its dynamics can be neglected. In this limiting case marginal stability is when $a_1 = 0$, i.e.

$$-k\,T_i\,\gamma = 2\mu_h/\omega_{nh}$$

$$\text{or} \quad -k\gamma = 2\mu_h/\omega_{nh}T_i \tag{8.4-4}$$

It is seen that only negative aberration slopes reduce stability margins; positive aberration slopes increase the effective damping. If $\mu_h = 1.0$, $\omega_{nh} = 10$, $T_i = 1.0$ and $k = 4$ then the negative slope must not exceed 0.05 or 5%. Homing heads are normally designed to be at least critically damped to help in this respect. The other extreme case is when the homing head and autopilot have equal undamped natural frequencies. The conditions for stability by the Routh-Hurwitz criterion are

$$a_1\,(a_2 - a_1 a_4/a_3) > a_3$$

With a more realistic model of the system it is seen from Fig 8.4-4 that both positive and negative aberration slopes can be harmful but that in general the negative slopes are much more critical. If both the homing head and autopilot are heavily damped the chance of radome aberration resulting in instability

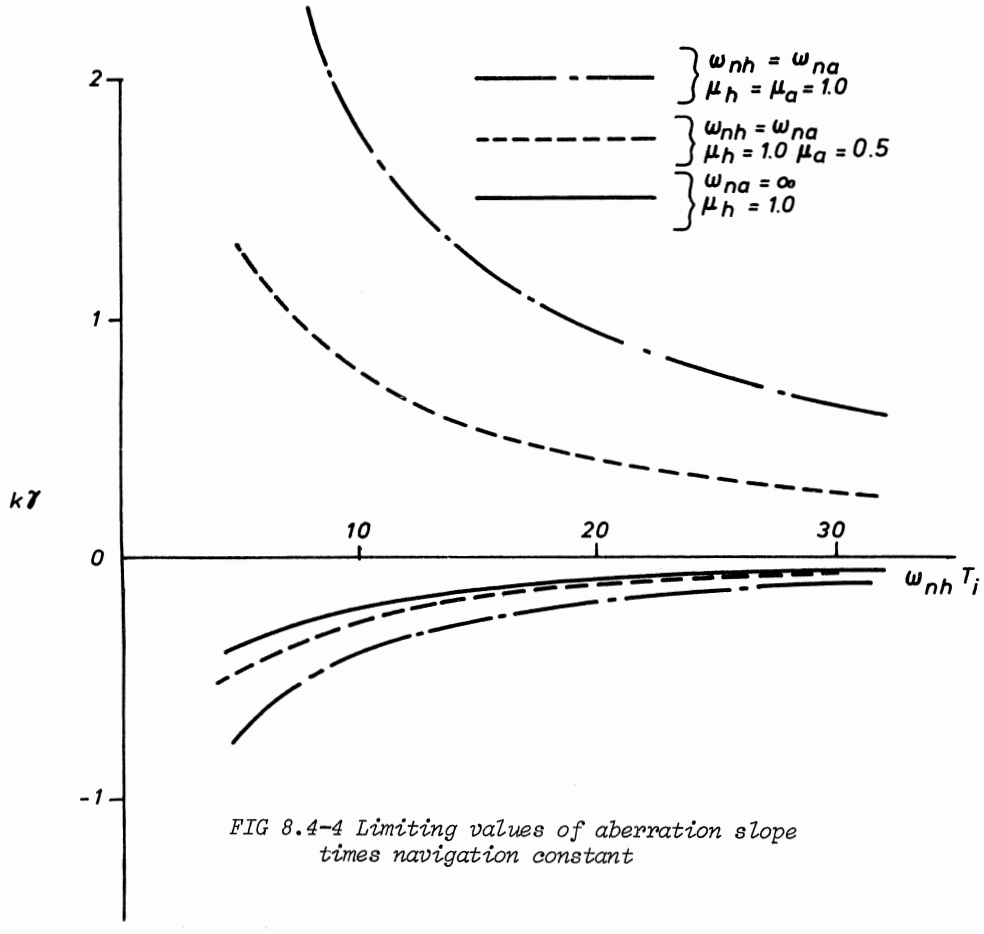


FIG 8.4-4 Limiting values of aberration slope times navigation constant

or near-instability is greatly reduced. In addition it is seen that the incidence lag is all-important. Typically this is about 0.5 secs or less for a supersonic missile flying at low altitudes but can easily be 2 secs or more at near-supersonic speeds and at altitude. If a missile operates over a range of speeds and/or heights and radome aberration is present, the homing head bandwidth has to be very low indeed say in the range 4-6 rad/sec. The reader is referred to a paper by Jackson (1) which deals more fully with this and other allied topics.

8.5 ISOLATED SIGHT LINE AND MISSILE COMPENSATION

If valve-controlled hydraulic jacks are used for the dish servo actuators the dish-body coupling is a maximum since a hydraulic servo is very stiff indeed

when the valve is in the mid position. It is a fair approximation to regard the servo as producing a dish movement relative to the body; the detector displacement is the servo displacement plus that due to the body. The coupled situation for a conventional rate gyro stabilised head is as shown in Fig 8.5-1.

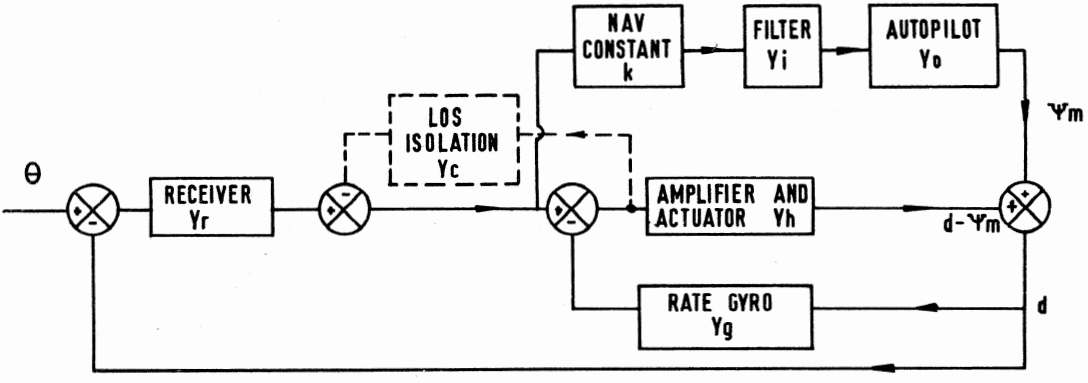


FIG 8.5-1 The use of line of sight isolation

where Y_r is the receiver transfer function plus any network considered necessary

Y_h is the servo amplifier and actuator t.f.

Y_g is the rate gyro t.f.

k is the navigation constant

Y_i is an inter-loop coupling filter t.f. (not always used)

Y_o is the autopilot t.f. for an output of body position

Considering the transfer function d/θ algebraic manipulation yields:

$$\frac{d}{\theta} = \frac{1 + kY_iY_o/Y_rY_h}{1 + Y_gY_r + 1\ Y_rY_h + kY_1Y_o/Y_h}$$

If one puts $k = 0$ then the transfer function is for the simple uncoupled head servo. The effect on the servo can be seen to be small if Y_h is very large; this is not surprising since a servo with a very large gain wipes out disturbances efficiently. For moderate values of Y_h (and/or large values of Y_o) the effect can be unstabilising. In practice additional feedback as shown called "line of sight isolation" is used. By choosing the transfer function Y_c carefully one can reduce the effect of changes in receiver axis from affecting measurements of sight line rate.

A second technique for improving the degree of system stability over and above that obtainable with isolated line of sight is to attempt to nullify the effect

of body motion by an explicit signal from the autopilot rate gyro (assuming there is one) as shown in Fig 8.5-2. This method is usually called "missile compensation".

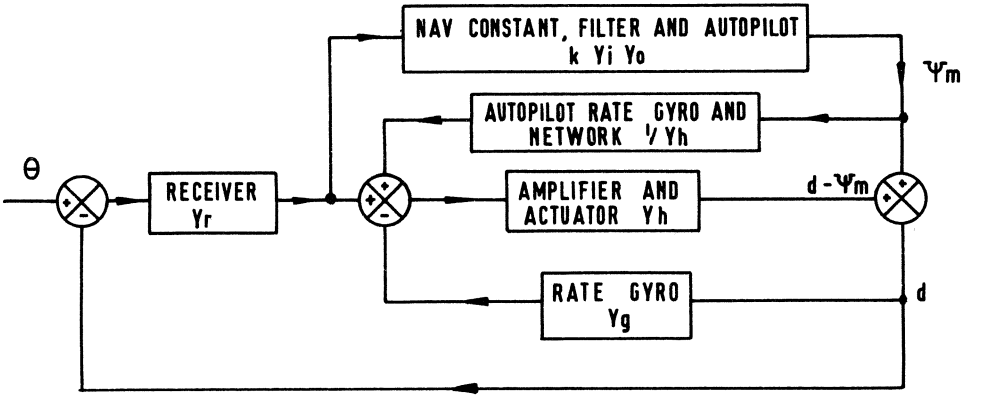


FIG 8.5-2 The use of missile compensation

If the rate gyro transfer function plus network can be made the inverse of the servo amplifier and actuator transfer function then complete cancellation occurs. If the actuator is non-linear then possibly 80% cancellation can be obtained.

REFERENCE

1. JACKSON R.F. The choice of autopilot bandwidth in a homing missile. British Aircraft Corporation report ST 7409, September 1972.

CHAPTER 9

PROPORTIONAL NAVIGATION AND HOMING GUIDANCE LOOPS

9.1 INTRODUCTION

We have seen that command systems are designed to maintain a near-constant guidance open loop gain. We have also seen that there are three points which are relevant to the guidance loop, the stationary tracker on the ground, the missile and the target. In homing systems the target tracker is in the missile and the ultimate source of the energy entering the missile homing head is irrelevant as far as any of the arguments or discussions in this chapter are concerned, i.e. we are concerned with the behaviour of the missile with respect to the target as a result of energy emanating from the target. Homing guidance is sometimes called "two-point guidance". It follows that as the missile approaches the target in homing systems the gain of the angular error detector in the missile increases inversely as $1/r$ where r is the instantaneous range of the missile to the target; a given lateral displacement off the tracker axis appears as an ever-increasing angle as the engagement proceeds. When considering homing guidance one must keep this simple fact constantly in mind as it is a source both of great strength and great weakness. Since there must be a number of lags in the loop it appears therefore that the system must eventually go unstable. However, it will be shown that two-point guidance allows one to use sight-line rate not sight-line angle as a guidance signal and hence one can regard the guidance loop as containing one inherent differentiation which will offset one of the inherent integrations. There are no immediate concerns over guidance loop stability at long ranges; but this will have to be examined in detail for short ranges.

9.2 A PARTICULAR CASE

In order to obtain some feel for the problems before resorting to mathematics it is worth considering the special case of an interception when missile and target speeds are constant and $U_m/U_t = 2$ say. The target is assumed to fly straight and we aim at the target in the same way as we would with a CLOS or beam-riding system. The guidance loop is engineered to produce a rate of change of trajectory ($\dot{\psi}_f$) which is k times the rate of change of sight line

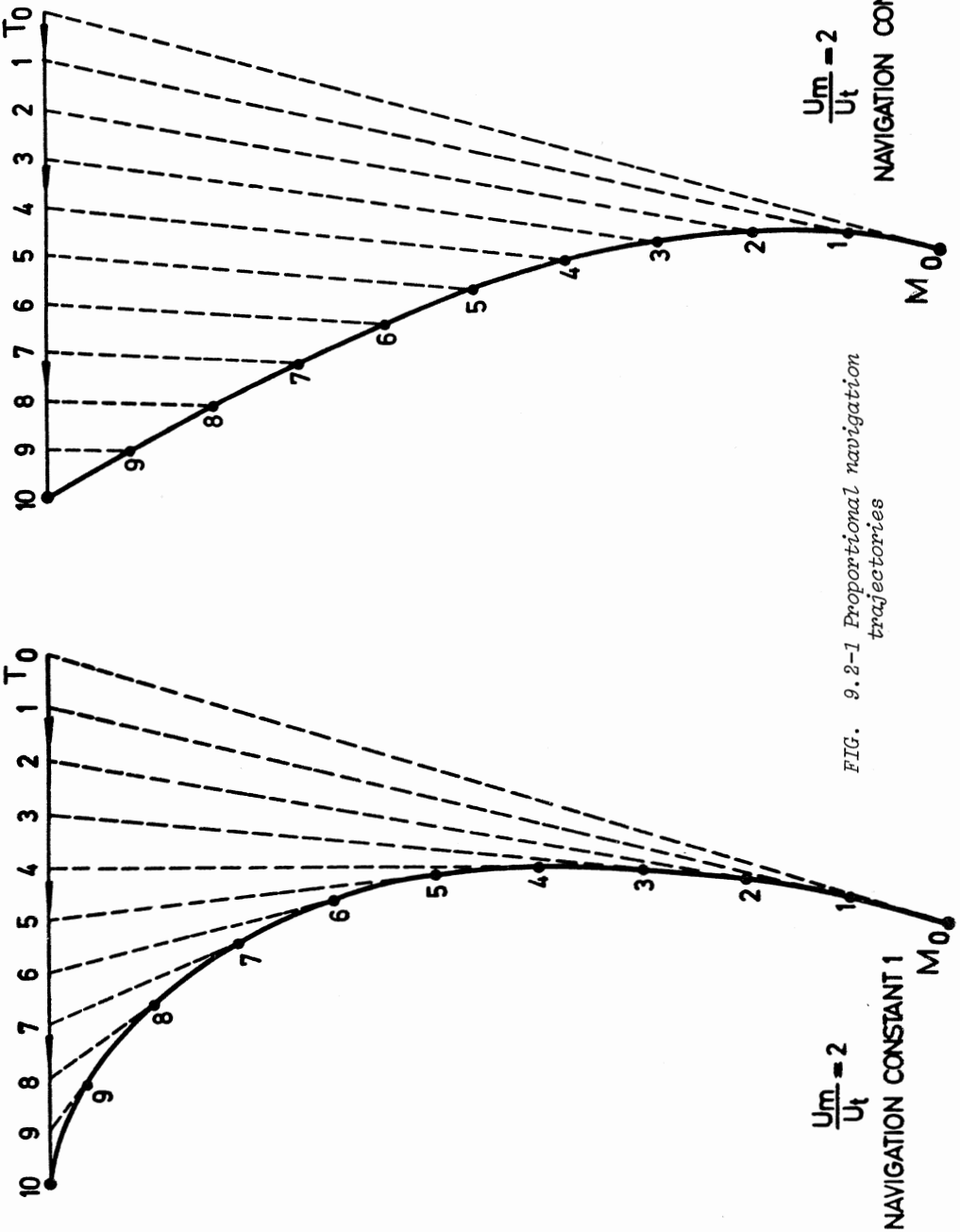


FIG. 9.2-1 Proportional navigation trajectories

($\dot{\theta}$) and k is called the navigation constant, i.e.

$$\dot{\psi}_f = k\dot{\theta} \tag{9.2-1}$$

Two cases are considered (a) $k = 1$ and (b) $k = 4$ and lags on the system are neglected. In Fig 9.2-1 M_o, M_1, M_2 and T_o, T_1, T_n are positions of the missile and target at launch and at successive intervals of time after launch. Dotted lines represent the sight line. If the navigation constant is unity then it is not difficult to see that as the trajectory direction changes at the same rate as the sight line and one aims at the target in the first place then the line drawn tangential to the missile flight path must start and remain coincident with the sight line. Since a tangent to the flight path indicates the instantaneous direction of the flight path such a trajectory is often referred to as a "pursuit course" as it is the sort of trajectory a dog might conceivably follow if chasing a rabbit. He always heads for the "target" and never attempts to aim ahead. If a navigation constant of say four is used, initially the sight line rate must be the same as in the first case, but the missile steering commands are four times as great; as a result the missile veers off much more to the left. Examination of the diagram shows that the sight line rate reduces as the engagement proceeds. It is most important to realise that such a guidance law automatically establishes a lead angle. Indeed, one can take the argument further. Consider missile and target flying straight at constant speeds, U_m and U_t respectively and intercepting at I, Fig 9.2-2. Since

$$\frac{M_oI}{T_oI} = \frac{U_m}{U_t} \tag{9.2-2}$$

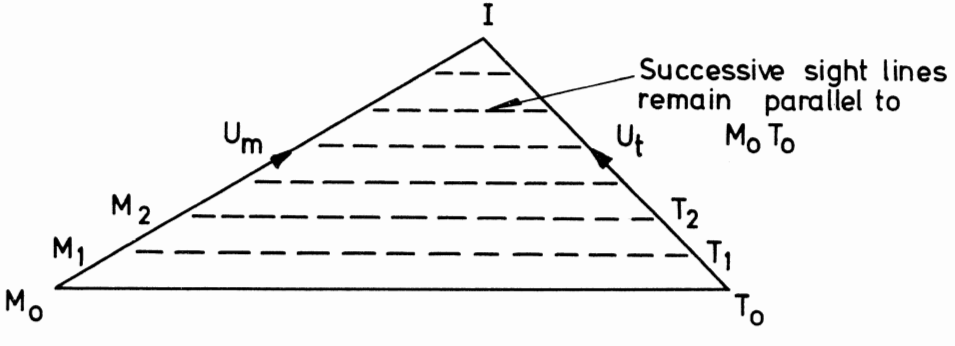


FIG. 9.2-2 A constant bearing collision course

then it is clear that the triangles M_oT_oI, M_1T_1I, etc are similar triangles and therefore successive sight lines M_oT_o, M_1T_1 etc are parallel i.e. the sight line does not rotate. This very simple fact is the basis of proportional navigation. If the sight line does not rotate in space then no steering commands are necessary as one is on a collision course. If the sight line does rotate (i.e. $\dot{\theta}$ exists) then a change of trajectory direction is required and it must be in such a sense as to reduce $\dot{\theta}$. Clearly the plane of the manoeuvre must be in the plane of $\dot{\theta}$. Imagine the missile to be roll position stabilised. The homing head measures the vertical component and the horizontal component of sight line rate and passes these signals suitably scaled to the elevator and rudder servos respectively. If the missile is not roll position stabilised no resolution is necessary as the homing head rotates with the missile and missile servos; so no vertical reference and resolver is necessary.

The concept of proportional navigation probably originated with mariners who must have known for centuries that if any object, moving or stationary, appears stationary and looms larger and larger then a collision is inevitable unless a change of course is made. It will be shown that in a well designed homing system lateral accelerations tend to decrease as one nears the target (providing the latter does not manoeuvre). In command systems for approaching targets this can never be so; lateral accelerations tend to increase towards impact. Consider a simplified situation as shown in Fig 9.2-3. Suppose the missile maximum lateral acceleration during an

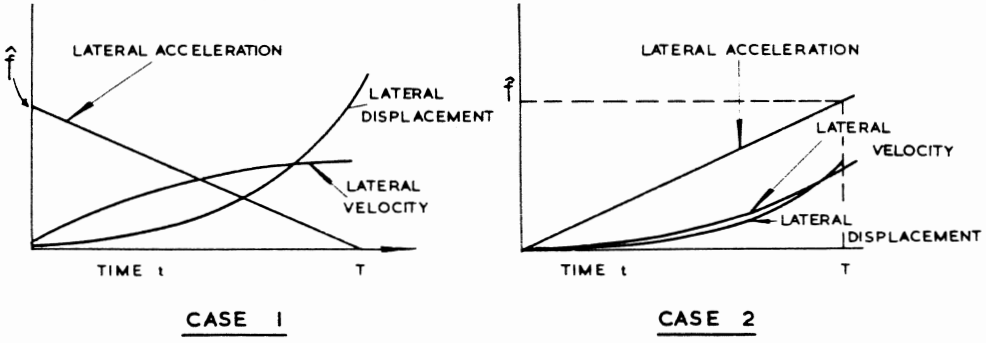


FIG. 9.2-3 Missile acceleration, velocity and displacement/time

engagement of duration T seconds is $\hat{f}$ in both cases, but assume it decreases linearly with time in case 1 and increases linearly with time in case 2. If no allowance is made for the fact that the direction of latax is continuously changing then:

	Case 1	Case 2
missile lateral acceleration	$f = \hat{f}\ (1 - t/T)$	$= \hat{f}\ t/T$
lateral velocity	$= \hat{f} \int (1 - t/T)dt + C_1$	$= \frac{\hat{f}}{T} \int t\ dt + C_2$
	$= \hat{f}\ (t - t^2/2T) + C_1$	$= \hat{f}\ t^2/2T + C_2$
lateral displacement	$= \hat{f}\ (\frac{t^2}{2} - \frac{t^3}{6T}) + C_1 t + C_3$	$= \frac{\hat{f} t^3}{6T} + C_2 t + C_4$

and for zero initial conditions, putting $t = T$ these expressions reduce to;

$$\frac{\hat{f}T^2}{3} \quad \text{and} \quad \frac{\hat{f}T^2}{6}$$

This is not surprising as lateral displacement is the moment of area of the lateral acceleration/time graph and the moment of area about the point $t = T$ for case 1 is twice that of case 2. Put in another way if a lateral displacement has to be made by the missile because the target velocity is not directed towards the missile launcher then a homing missile using proportional navigation will require less latax than one using CLOS or beam-riding guidance. Of course, in homing systems given some computing facilities one can aim off in the first place to reduce latax requirements still further. We now proceed to a more general analysis.

9.3 THE MATHEMATICAL MODEL

There is a formidable difficulty in arriving at a general solution of the equations describing the behaviour of homing systems using proportional navigation. One can assume a linear model for the homing head and autopilot and linearity of the kinematics can be obtained by assuming that the effect of target manoeuvre, noise, or a heading error of the missile from the ideal direction which would lead to a constant bearing collision course, as a small perturbation from that collision course. This allows one to make small angle approximations. If one's attention is directed to the closing stages of the engagement this could be argued to be a fair approximation as the missile is usually on a near-constant bearing trajectory just before impact. The assumption is less valid if the target makes a large evasive manoeuvre during this period as our simplified geometry could undergo some distortion. Fig 9.3-1 defines the interception geometry.

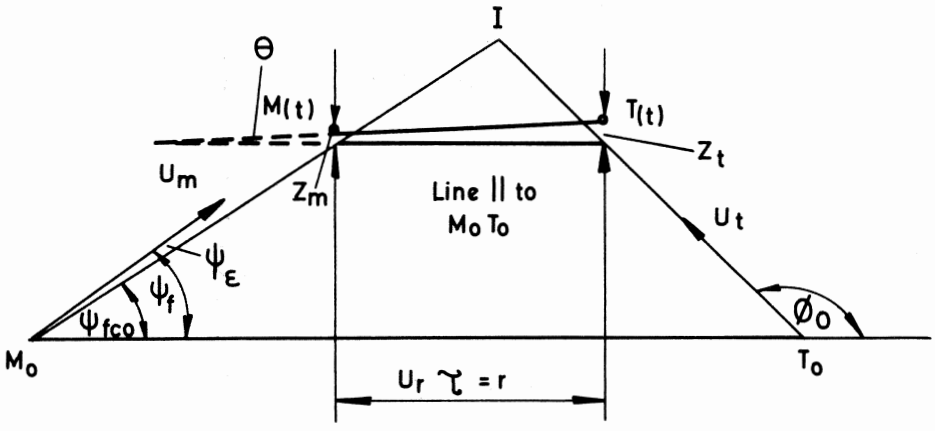


FIG. 9.3-1 Interception geometry

The target is flying at ϕ_o to the original sight line M_oT_o which is regarded as a reference direction, and the missile is flying at a small angle ψ_e to the correct flight path to obtain a collision at I. The correct flight path angle ψ_{fco} is given by:

$$U_m \sin \psi_{fco} = U_t \sin \phi_o \tag{9.3-1}$$

There is in fact an imaginary line which runs up the paper parallel to M_oT_o at a velocity of $U_m \sin \psi_{fco}$. Any perturbation of the missile and target perpendicular to this line is denoted by z_m and z_t respectively. The following relationships follow:

$$\tan \theta = (z_t - z_m)/r = (z_t - z_m)/U_r\tau \tag{9.3-2}$$

and for small angles

$$\theta \simeq \tan \theta \tag{9.3-3}$$

where θ = sight line angle

U_r = the relative or closing velocity = $U_m \cos \psi_f - U_t \cos \phi_o$

and τ = the time to go considered zero when $r = 0$.

The method used by the author to compute miss distance in one plane is to compute $z_t - z_m$ at $r = 0$. This is not the true vector miss distance but it can be shown that this is very nearly so provided that the concept of small perturbations still holds.

How does such a model look in the conventional closed loop representation? If ψ_e is small we can regard $\psi_f \simeq \psi_{fco}$ because

$$\psi_f = \psi_{fco} \pm \psi_e \tag{9.3-4}$$

and $\cos \psi_f = \cos (\psi_{fco} \pm \psi_e) = \cos \psi_{fco} \cos \psi_e \pm \sin \psi_{fco} \sin \psi_e$ (9.3-5)

It is assumed that the homing head steady state gain is k_1 volts rad^{-1} sec and can be modelled dynamically as a second order lag. This signal is passed to a navigation constant amplifier of gain k_2 volts/volt. The autopilot employs accelerometer plus rate gyro feedback so the steady state gain is sensibly constant at k_3 m sec^{-2} volt^{-1}, and again is adequately modelled as a second order lag. The combined gain of these three items has been shown as $K = k_1 k_2 k_3$. We observe that only relative movements perpendicular to the sight line will be seen by the error detector in the homing head. The closed loop model therefore is as shown in Fig 9.3-2.

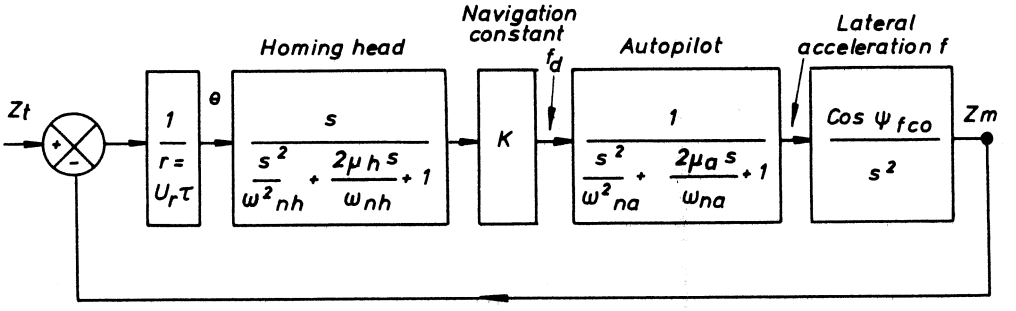


FIG. 9.3-2 Small perturbation closed loop representation of homing system

Since the time to go $\tau = T - t$, where T is the time to go for the initial positions $M_o T_o$, is common to all systems and engagements, all systems will, for the same dynamic lags, be identical if

$$\frac{K}{U_r} \cos \psi_{fco} = \text{a constant} = \alpha, \text{ say.} \qquad (9.3\text{-}6)$$

This constant α has no dimensions and is usually known as the kinematic gain or kinematic stiffness; it is not the open loop gain. The open loop gain is α/τ. If an optimum value for α exists then it follows that one should adopt a navigation constant such that

$$K = \frac{U_r \alpha}{\cos \psi_{fco}} \qquad (9.3\text{-}7)$$

It will be shown that in most cases α should lie in the range 3.5 to 4. Now, ψ_{fco} cannot exceed about 45^o because of mechanical angle of look considerations and in most cases is likely to be much less than this. We can deal with this variable in real systems in one of three ways:

(a) neglect it altogether as is the usual case in CLOS and beam-riding systems. Its existence results in a small loss of open loop gain. OR

(b) make the approximation that $0.707 < \cos \psi_{fco} < 1$ and therefore

$\cos \psi_{fco} \simeq .85$. OR

(c) compute ψ_f or ψ_{fco}. The author knows of no system where this has been done. The computation or otherwise of U_r will be discussed later.

Suppose now that the missile autopilot uses rate gyro feedback only, and that the instrument feedback is lagged by a time constant which approximates to the incidence lag; such an autopilot produces a flight path rate as a result of a demand. For such a system denote the navigation constant as k (dimensions are rate of change of flight path/rate of change of sight line i.e. no dimensions) and since

$$f_y = \dot{\psi}_f U_m \tag{9.3-8}$$

then this system is equivalent to one with a lateral acceleration autopilot if

$$k = K/U_m \tag{9.3-9}$$

$$\text{i.e. } k = \frac{\alpha U_r}{U_m \cos \psi_{fco}} \tag{9.3-10}$$

It is perhaps easier to envisage such a system rather than the acceleration system. Suppose $U_m/U_t = 2$. Then:

(a) for head on attack $U_r/U_m = 1.5$ and $\cos \psi_{fco} = 1$

(b) for broadside attack $U_r/U_m = 0.866$ and $\cos \psi_{fco} = .866$ and

(c) for a tail chase $U_r/U_m = 0.5$ and $\cos \psi_{fco} = 1$.

For case (a) therefore one should demand a flight path rate six times the sight line rate for α to be 4. For case (b) the demand should be four times the sight line rate and for case (c) the demand for flight path rate should be reduced to twice the sight line rate.

There is a simple physical explanation why the navigation constant should be proportional to the relative velocity. In any given system the gain of the homing head increases inversely with range to go; and this is the same thing as saying as inversely with time to go times relative velocity. But in all situations there must be four seconds, three seconds, two seconds, one second to go etc. So a situation where the relative velocity is low e.g. a tail chase, results in a large gain for a long time due to the range effect; therefore the navigation constant should be reduced. When there is a head-on attack, a short range and therefore high gain, exists for a short time only and therefore the navigation constant should be set higher. We are dealing with a situation where the open loop gain is time varying, and all systems have the same gain at the same time to go if α is a constant. If $\alpha = 4$, then

the open loop gain is unity for $\tau = 4$, it is 2 for $\tau = 2$, it is 4 for $\tau = 1$ and 8 for $\tau = 0.5$.

9.4 A SUMMARY OF PREVIOUS WORK

All published works known to the author have made use of the simplifying assumptions of small perturbations about a constant bearing trajectory. Cornford and Bain (1) obtain a general solution for the trajectory of a missile. Due to the complexity of a general solution systems represented by a single time constant only are considered. Nevertheless this fundamental work established the significance of the kinematic gain α and all subsequent works in this field have pointed out that it is this quantity which is fundamental and not the navigation constant itself. Bain and Trebble (2) extend the analysis to a system represented dynamically by a quadratic lag and obtain general solutions for miss distance due to a target evasive manoeuvre, linear noise and angular noise. Discreet values of α of 1, 2 and 3 are considered. Bain (3) extends this work and includes an analysis of miss distance due to an initial missile heading error. Jenkins (4, 5) considers the miss distance due to a missile heading error, a target evasive manoeuvre and a target weave. Missile latax limits are placed on the missile for the first two cases. Again, the system is represented by a second order lag. Heap (6) produces some interesting trajectories for noise-free systems and discusses, in general terms, the effect of noise and optimum filters.

9.5 THE EFFECT OF A MISSILE HEADING ERROR

It is instructive to obtain a general solution for the lateral acceleration of the missile if it has a small initial aiming (i.e. heading) error ψ_e. A simple solution is available if all the lags in the system are neglected! This may seem to be too unrealistic to be of any value; nevertheless the reader will find that the solution is significant and provides useful information concerning the behaviour of homing systems when the engagement time is long compared with the system reaction time. The initial missile velocity perpendicular to the collision course is $U_m \psi_e$ and the component perpendicular to the initial sight line $M_o T_o$ is $U_m \psi_e \cos \psi_{fco}$. Now $\dot{z}_m$ is this initial velocity plus or minus any additional lateral velocity as a result of steering commands. Also $z_t = \dot{z}_t = \ddot{z}_t = 0$ for all values of t. The closed loop diagram is shown in Fig 9.5-1. Note that $U_m \psi_e \cos \psi_{fco}$ can be regarded as an input to the system; in fact the system would detect no difference in the

situation if the initial heading was correct and at $t = +0$ the target instantaneously changed its course to produce a velocity perpendicular to the original sight line $M_o T_o$ equal in magnitude to $U_m \psi_e \cos \psi_{fco}$. For this reason a heading error is often referred to as a step velocity input.

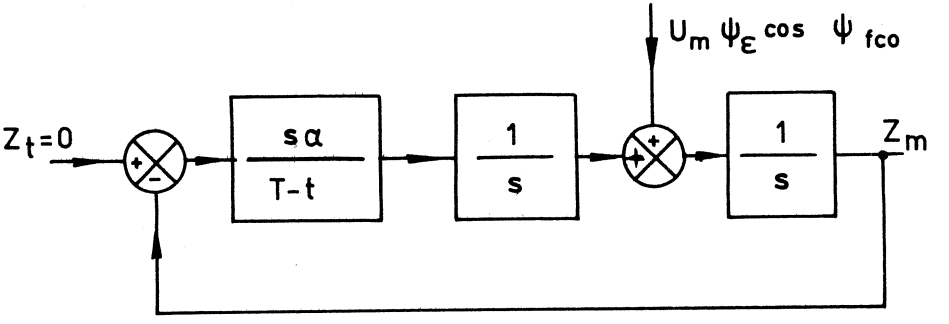


FIG. 9.5-1 Lag-free system with step velocity input

By inspection, the output z_m due to the input can be written down thus:

$$\frac{z_m}{U_m \psi_e \cos \psi_{fco}} = \frac{1}{s + \frac{\alpha}{T - t}} \tag{9.5-1}$$

or, in differential equation form:

$$(D + \frac{\alpha}{T - t}) z_m = U_m \psi_e \cos \psi_{fco} \tag{9.5-2}$$

The solution to this equation is:

$$z_m = U_m \psi_e \cos \psi_{fco} \frac{(T - t)}{(\alpha - 1)} \{1 - (1 - t/T)^{\alpha-1}\} \tag{9.5-3}$$

This can be differentiated twice to give $\ddot{z}_m$:

$$\ddot{z}_m = - U_m \psi_e \cos \psi_{fco} \frac{\alpha}{T} (1 - t/T)^{\alpha-2} \tag{9.5-4}$$

and since $\ddot{z}_m = f_m \cos \psi_{fco}$ this can be rearranged non-dimensionally thus:

$$- \frac{f_m T}{U_m \psi_e} = \alpha(1 - t/T)^{\alpha-2} \tag{9.5-5}$$

For $\alpha = 1, 2, 3, 4$ equation (9.5-5) can be shown in Fig 9.5-2:

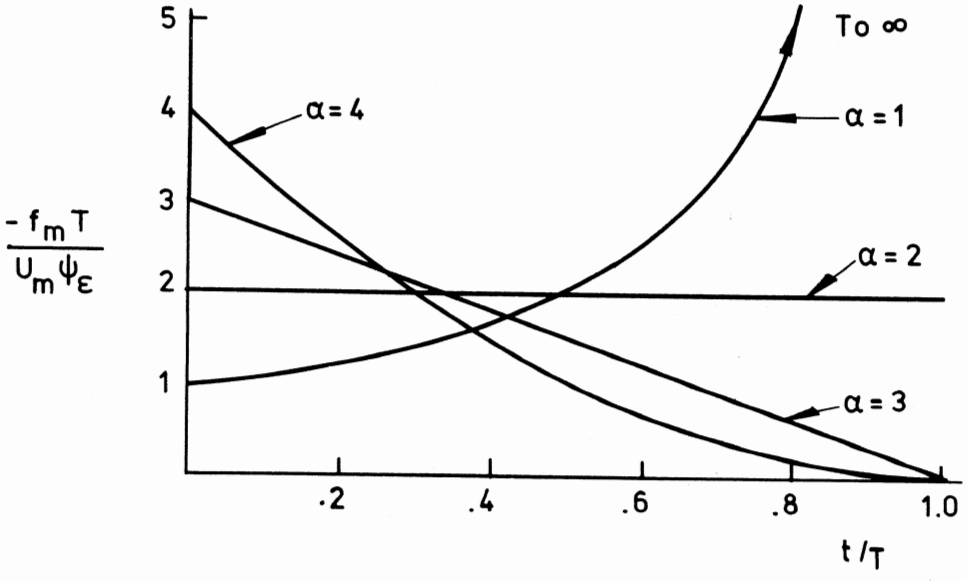


FIG. 9.5-2 Missile latax for a step velocity input-lag free system

It is seen that if $\alpha = 3$, the missile lateral acceleration decreases linearly with time and if $\alpha = 4$, the initial value is greater of course but the missile lateral acceleration towards the end of the engagement is very small; the trajectory approximates to a constant bearing course eventually. When $\alpha = 2$ the lateral acceleration remains constant i.e. the trajectory is an arc of a circle. The validity of this solution can easily be checked. The value of z_m at $t = T$ due to a heading error with no steering commands is:

$$z_m = U_m \psi_e \cos \psi_{fco} T \tag{9.5-6}$$

If the missile lateral acceleration f_m = a constant, then the lateral displacement perpendicular to the initial sight line at $t = T$ is given by,

$$z_m = \tfrac{1}{2} f_m \cos \psi_{fco} T^2 \tag{9.5-7}$$

The value of f_m is given by substituting $\alpha = 2$ in equation (9.5-5) and is:

$$f_m = -2 U_m \psi_e / T \tag{9.5-8}$$

Substituting this value in equation (9.5-7) yields:

$$z_m = - U_m \psi_e \cos \psi_{fco} T \tag{9.5-9}$$

Thus the total displacement of the missile at $t = T$ is zero. Inspection of equation 9.5-5 shows that in general lateral acceleration is inversely proportional to the time of the engagement and this clearly indicates that missile lateral acceleration limits may restrict the system coverage for short

range engagements; in this one sense there is a similarity between command systems and homing systems.

What is the optimum value of α? Suppose we wish to minimise the integral with respect to time of the missile squared lateral acceleration with α the variable. Using equation (9.5-5) we find

$$\int_0^T f_m^2 \, dt = \frac{U_m^2 \psi_e^2}{T} \left(\frac{\alpha^2}{2\alpha - 3}\right) \qquad (9.5\text{-}10)$$

$$\text{For optimum } \alpha \quad \frac{\partial}{\partial \alpha} \int_0^T f_m^2 \, dt = 0 \text{ from which } \alpha = 3$$

Before looking at results pertaining to a more realistic model of the system we should consider what would happen if α is less than 2. Inspection of equation 9.5-5 shows that with α less than 2 the demand for lateral acceleration tends to infinity just before interception and therefore the legitimate steering commands cannot be implemented at this crucial stage of the interception. A miss must result and this could be large. If α is set high (5 say) a large lateral acceleration will be demanded in the early stages and the missile may stay on limits for a while. All this means is that the lateral acceleration will not decay so rapidly. In this respect homing systems are completely different from command systems. If the missile limits on acceleration for more than a very short period in the latter case it will rapidly move out of the line of sight or beam and this means an abortive mission. Homing systems are far more flexible.

Now to results pertaining to a more realistic model of a system. Figs 9.5-3 to 9.5-10 show missile lateral acceleration/time for a heading error. The results have been obtained by the author on an ICL 1903A digital computer using an analogue simulation program. It is assumed that the homing head is the main lag in the system and that both homing head and autopilot can be represented by second order lags. If this is accepted then all systems must lie between the two extremes of

(a) $\omega_{na} = \omega_{nh}$ (autopilot and homing head lags equal)

(b) $\omega_{na} = \infty$ (autopilot infinitely faster than homing head and therefore effectively only one quadratic lag in the system).

Results have been shown for a badly and a well damped homing head, for practical values of α of 2.5 and 4.5 these being on the low and high sides respectively and for relatively "short" and "long" engagements defined by $\omega_{nh}T$ = 10 and 30 respectively. For comparison the results for a lag free system are also included. It will be seen that short engagements, inadequate

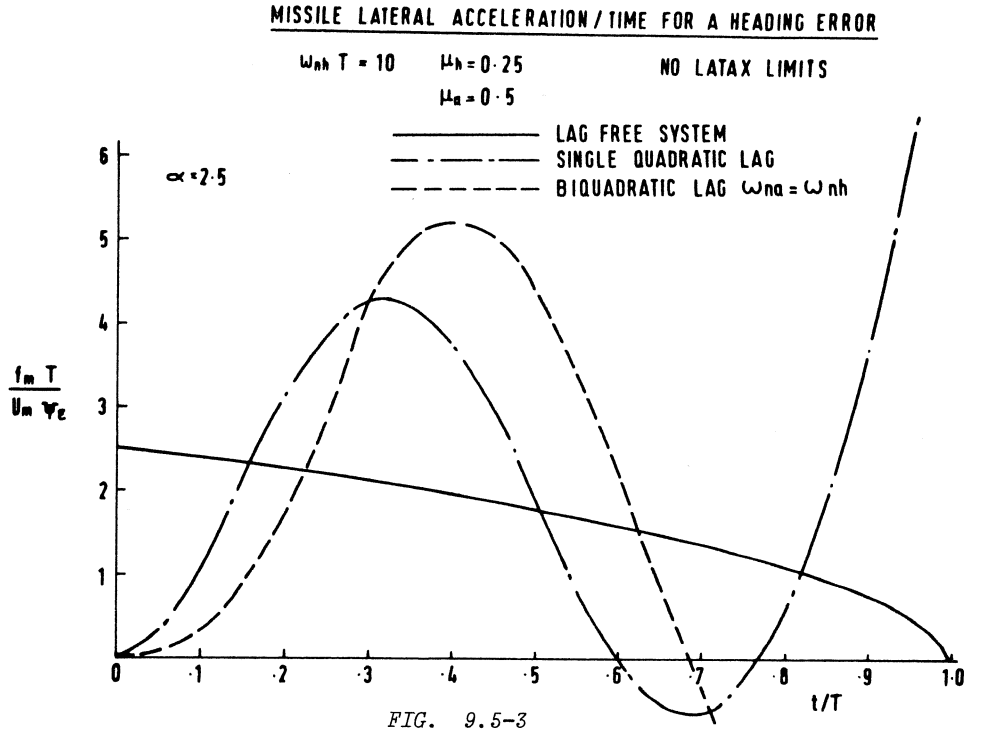


FIG. 9.5-3

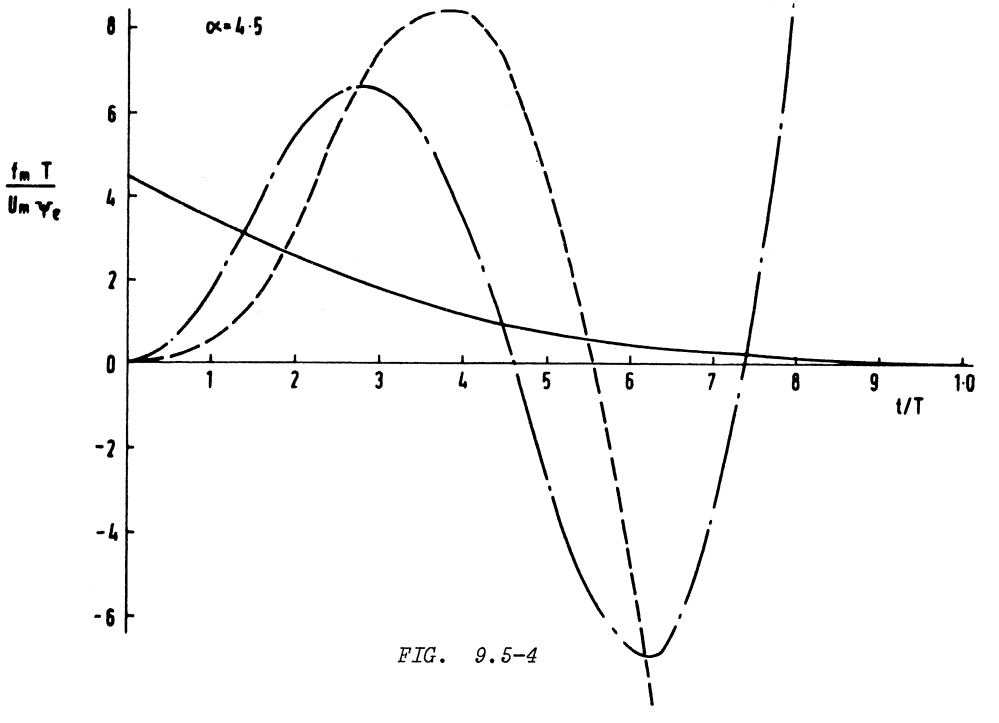


FIG. 9.5-4

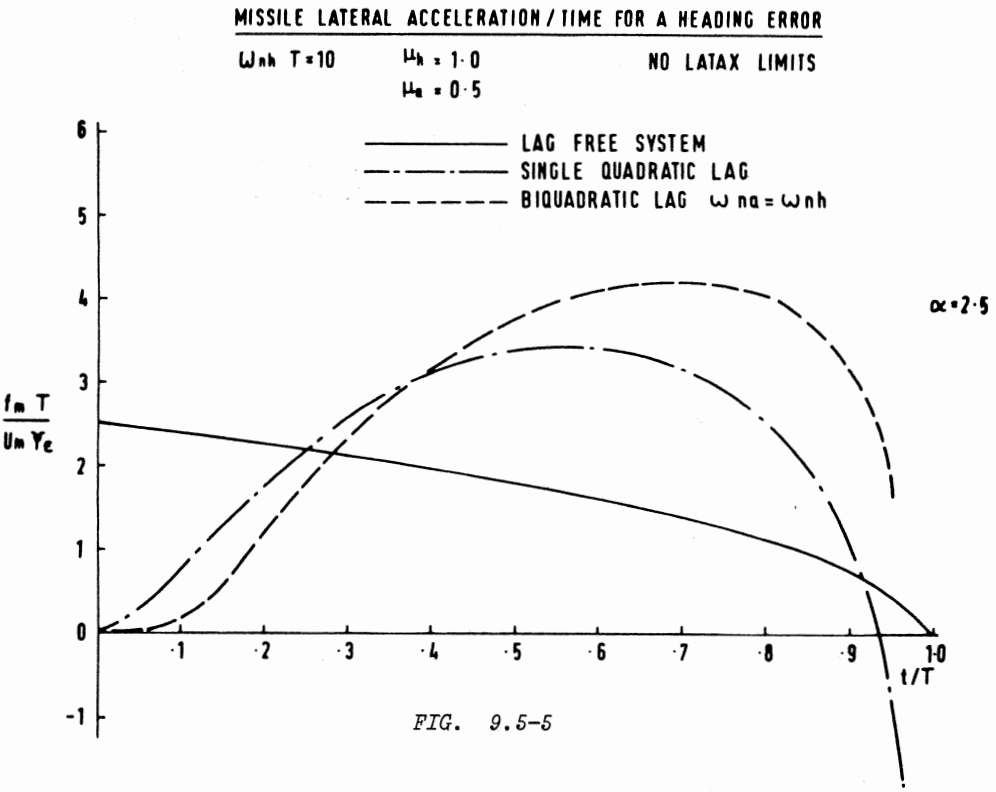


FIG. 9.5-5

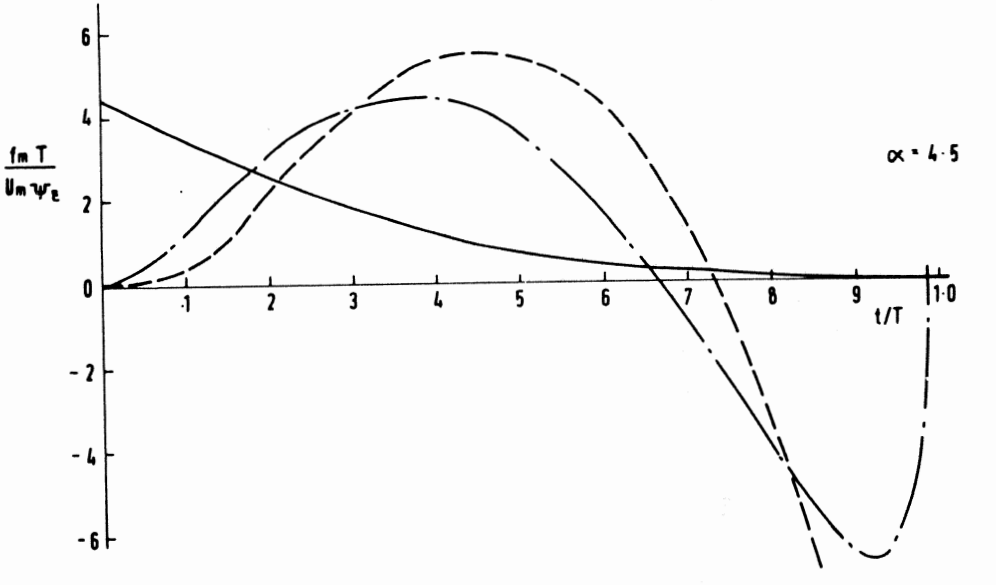


FIG. 9.5-6

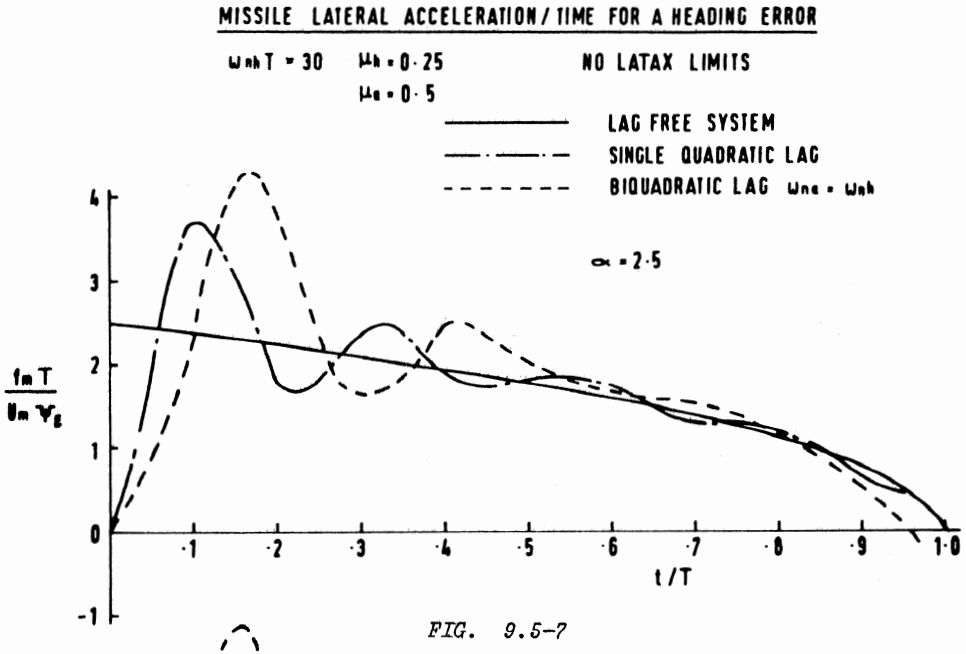


FIG. 9.5-7

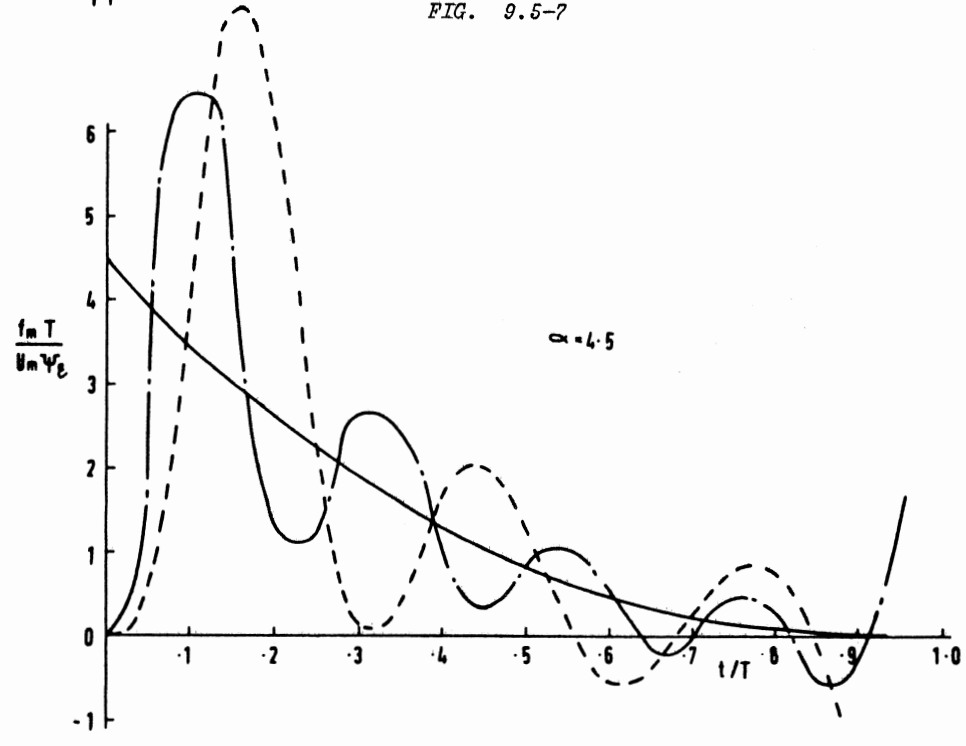


FIG. 9.5-8

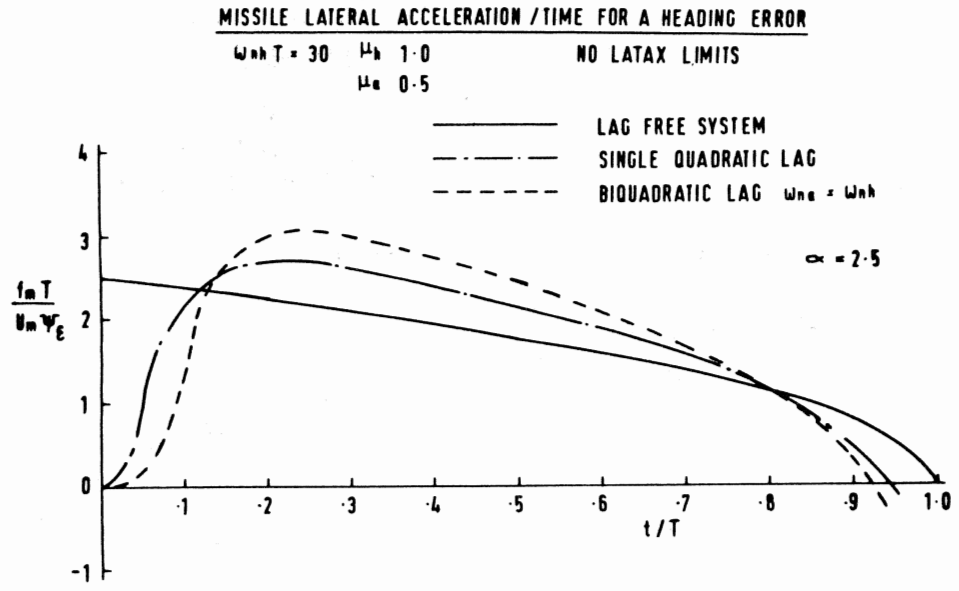


FIG. 9.5-9

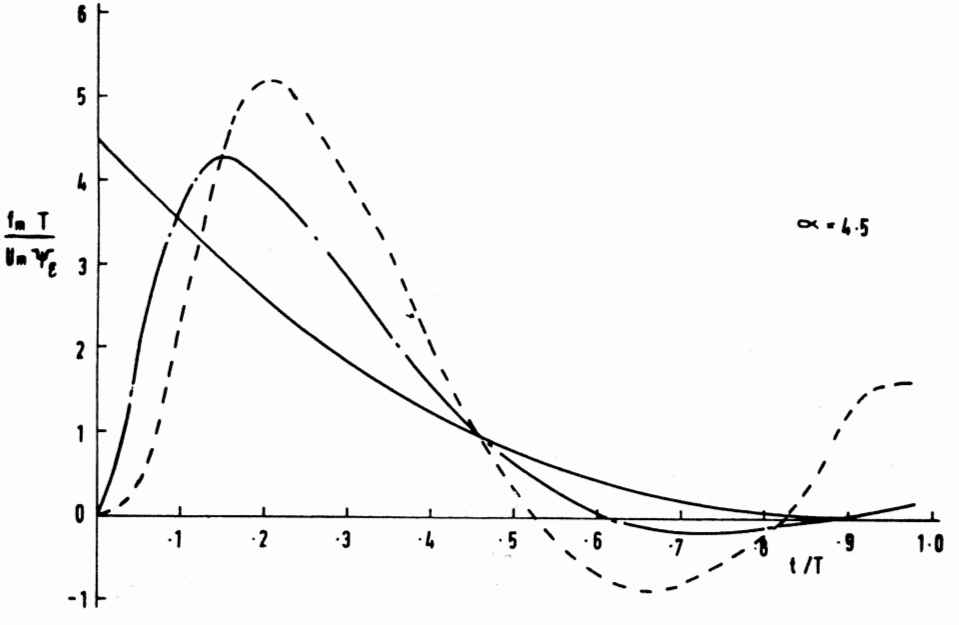


FIG. 9.5-10

damping and high values of α result in oscillatory responses. The autopilot damping ratio has been set to 0.5 in all cases.

Figs 9.5-11 to 9.5-16 show graphs of miss distance M_h due to a heading error. These graphs have been plotted against a normalised length of engagement $\omega_{nh}T$; all points on all graphs have been plotted for $t = T$. For α = 2.5 to 3.5 and reasonable damping it is seen that there should be a very small miss if $\omega_{na}T > 20$ for equal dynamic lags, and if $\omega_{nh}T > 15$ if the autopilot is much faster than the homing head. Nevertheless with α = 4.5 and poor damping of the dominant lag the system takes a long time to settle and this effect is more pronounced when the lags are equal. On the whole damping which is on the high side is preferable to light damping but the response can be rather sluggish with well damped systems and low values of α. The general conclusion is that a proportional navigation law will tend to produce a zero miss due to a heading error provided there is time for the transient to decay and this time depends on the nature of the lags in the system and the value of α. For equal lags, α = 3.5 and μ_h = 1.0 the maximum miss will occur if $\omega_{nh}T \simeq 3$ and is given by

$$\frac{M_h\ \omega_{nh}}{U_m\ \psi_e} \simeq 2.7, \text{ from fig 9.5-15}$$

If U_m = 500, ψ_e = 0.3 and ω_{nh} = 6 then M_h = 67.5m. Since the time of engagement is only 3/6 = 0.5 secs it follows that the miss would have been 500 x 0.3 x 0.5 = 75m if there had been no reaction at all. These graphs illustrate why an assumption of "at least three seconds homing time is required" is often made in preliminary studies on the coverage of homing systems. If missile homing head and autopilot bandwidths of 6 and 12 rad/sec respectively are adopted then three seconds homing time means that the normalised engagement time is given by $\omega_{nh}T$ = 18 and this is just about a minimum engagement time for such a system. A faster system needs less homing time of course.

Consider now the possible effects of the system becoming unstable at a short range to go when the open loop gain is high. If instability sets in say half a second before nominal impact, the control fins go hard over instantaneously and 15g is developed 0.25 secs later, then the miss due to this effect is 0.5 x 15 x 9.81/16 = 4.6 m. The author considers that most well designed systems go unstable rather later than this and the effect of instability on accuracy is usually very small. To convince the reader that homing systems can easily go unstable in practice is, oddly enough, not easy. In the

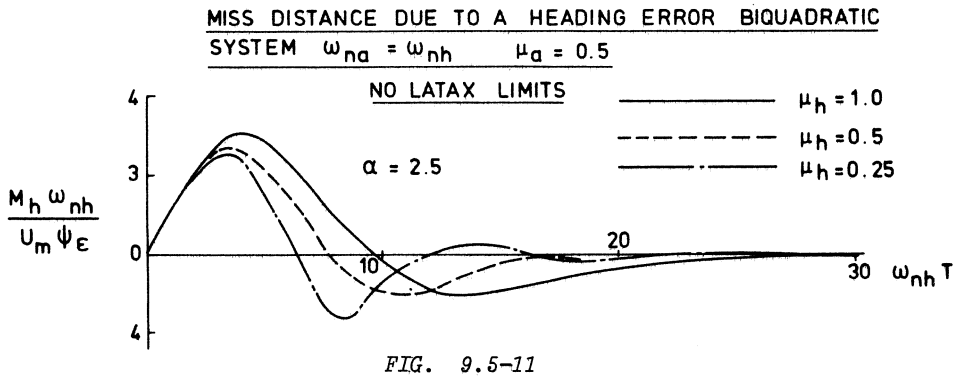


FIG. 9.5-11

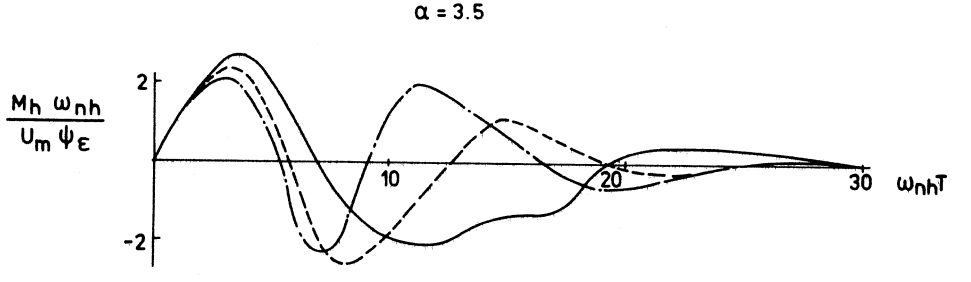


FIG. 9.5-12

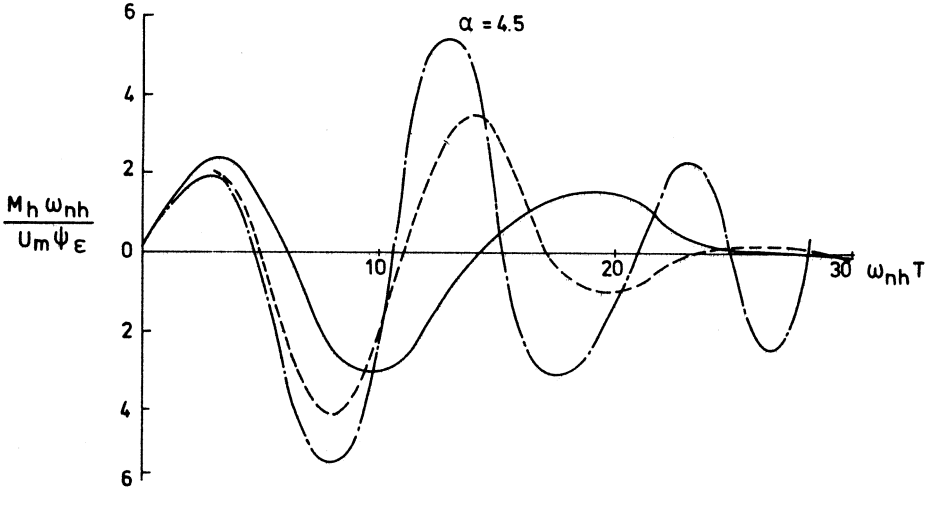


FIG. 9.5-13

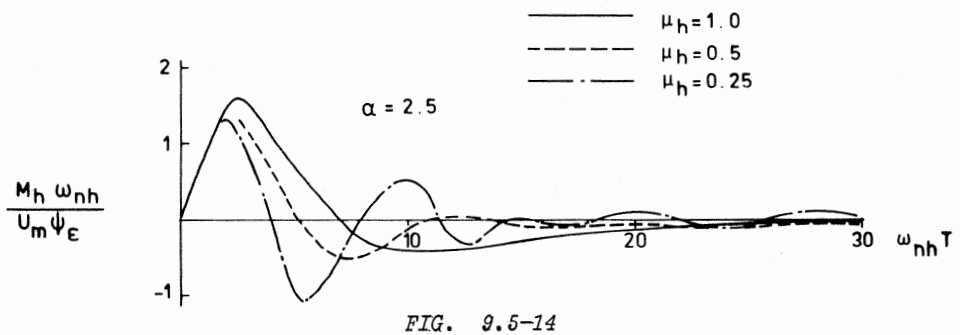


FIG. 9.5-14

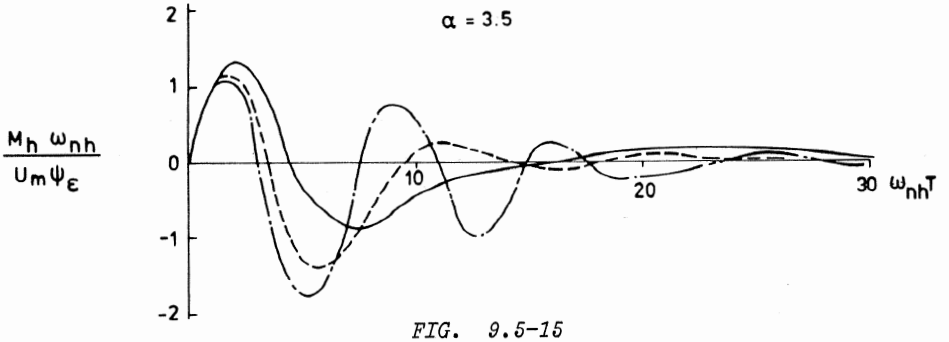


FIG. 9.5-15

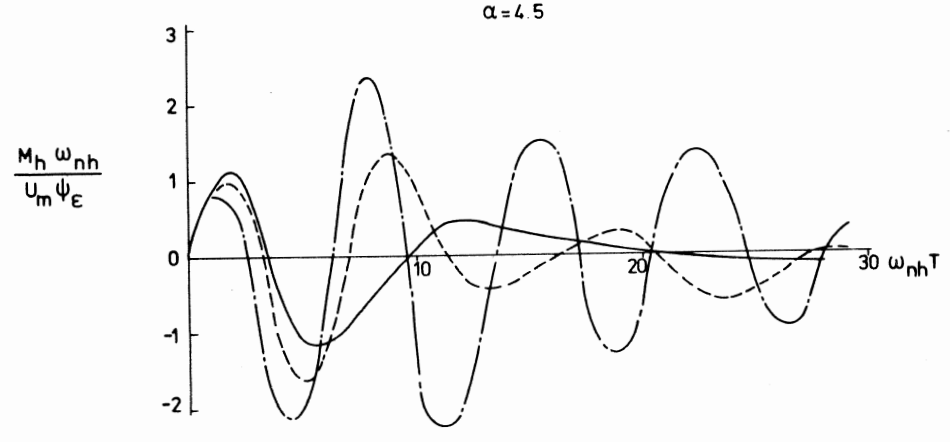


FIG. 9.5-16

absence of noise (no real system is ever entirely free from noise) a long engagement results in the transient decaying before instability sets in, and the effect is not apparent. If the engagement is very short oscillations do not have enough time to build up. Consider Figs 9.5-17 to 9.5-19 which have been computed for $\omega_{nh}T = 20$ and an underdamped homing head; this engagement is neither "short" nor "long". The response of a lag-free system is shown for comparison. Any response of a real system which diverges from this with time can be regarded as unstable. It is not possible to discern instability for $\alpha = 2.5$, but for $\alpha = 3.5$ a system represented by a single quadratic lag could be argued to be unstable later than $t/T = 0.75$ say. For $\alpha = 4.5$ instability sets in at about $t/T = 0.6$. By inspection of Fig 9.3-2 we can write down the transfer function for a system with an infinitely fast autopilot (i.e. only one quadratic lag effectively):

$$\frac{z_m}{z_t} = \frac{1}{(s^2/\omega_{nh}^2 + 2\mu_h s/\omega_{nh} + 1)\ \tau_s/\alpha + 1} \qquad (9.5\text{-}11)$$

It is realised that the Routh-Hurwitz criterion of stability applies to polynomials with constant co-efficients and that there are other stability criteria which are applicable here. Nevertheless what would this criterion predict if (erroneously) the coefficients were "frozen" at a given value of τ? Such a system is unstable when $a_1 a_2 < a_o a_3$ i.e.

$$\frac{\tau}{\alpha} \times \frac{2\mu_h}{\omega_{nh}} < \frac{1}{\omega_{nh}^2}$$

i.e. $$\tau < \alpha/2\mu_h\ \omega_{nh}$$

i.e. $$\tau/T < \alpha/2\mu_h\ \omega_{nh}T \qquad (9.5\text{-}12)$$

Figs 9.5-17 to 9.5-19 have been computed for $\omega_{nh}T = 20$ and $\mu_h = 0.25$. Therefore in this context equation 9.5-12 can be simplified to:

$$\tau/T < \alpha/10 \qquad (9.5\text{-}13)$$

Since $\tau = T - t$ and $\tau/T = 1 - t/T$ it is seen that for $\alpha = 3.5$ instability is predicted to set in at $t/T = 0.65$, and for $\alpha = 4.5$ at $t/T = 0.55$; and the agreement appears to be good. It is very easy to demonstrate early instability for systems with high kinematic gains of 5 or more.

Nevertheless, even though the Routh-Hurwitz criterion appears to give an approximate guide to instability in such a case it is pointed out that the homogenous equation for z_m obtained from equation 9.5-11 i.e.

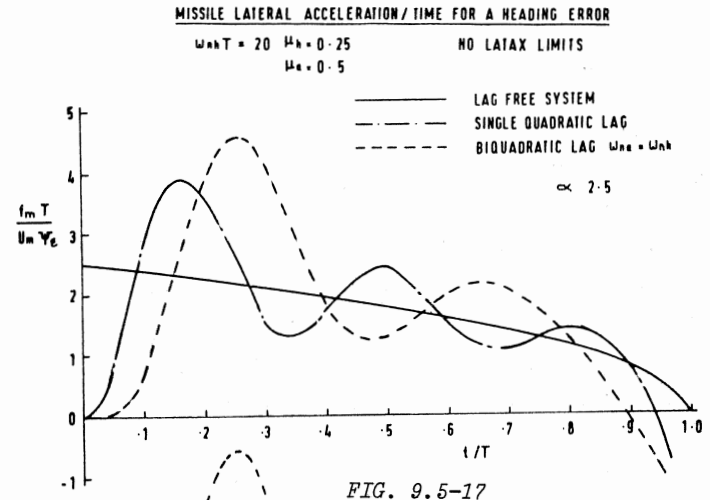


FIG. 9.5-17

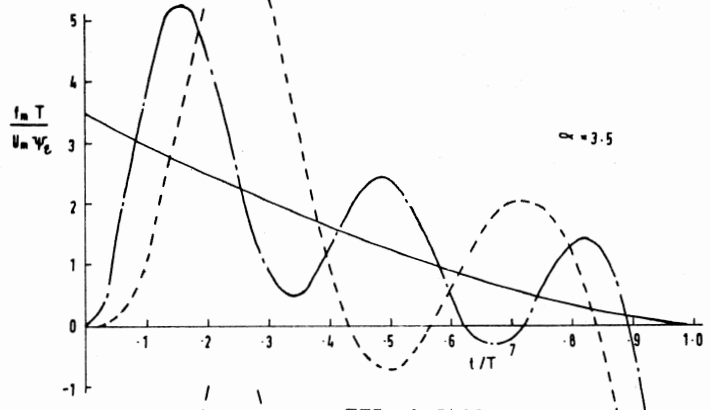


FIG. 9.5-18

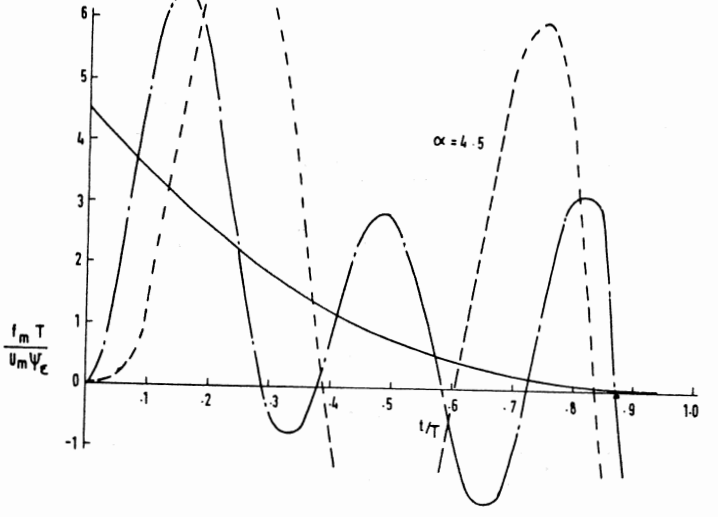


FIG. 9.5-19

$$\{\left(\frac{D^2}{\omega_{nh}^2} + \frac{2\mu_h D}{\omega_{nh}} + 1\right) \frac{(T - t)D}{\alpha} + 1\} z_m = 0 \qquad (9.5\text{-}14)$$

differs from that for $\dot{z}_m$ and $\ddot{z}_m$ due to time appearing in the equation i.e. there are different criteria for the instability of displacement, velocity and acceleration! Cornford and Bain (1) pointed this out and virtually dismissed the subject of instability. As far as is known the possibility of homing systems going unstable has not been discussed since.

The effect of missile latax limits is now discussed. Since for $\alpha > 2$ maximum demands occur early rather than late, if limiting occurs the missile will stay on limits for a while but may come off limits later and a hit may result. Consider therefore a system with α set to 3 and the "g" limiters set to a hypothetical but convenient value of $2\ U_m\ \psi_e/T$. The missile will try initially to accelerate laterally to $3\ U_m\ \psi_e/T$, see equation (9.5-5) and being unable to do so will remain on limits for the whole of the engagement. Since in any real system there will be a time lag before full latax is developed it follows that a miss must result. Hence, for an engagement to be successful α must always be set to > 2 and the "g" limiters must be set to a value in excess of $2\ U_m\ \psi_e/T$. A number of computer runs have been done with this in mind: in each successive run the only parameter that has been altered has been the value set on the limiters. Each series of runs was terminated when, as a result of prolonged limiting, the miss distance was seen to increase appreciably. As can be appreciated the longer the time of the engagement the more nearly does the critical value of limits approach $2\ U_m\ \psi_e/T$. If the demanded lateral acceleration f_{yd} is given by:

$$f_{yd} = c\ U_m\ \psi_e/T \qquad (9.5\text{-}15)$$

then Table 9.5-1 gives some values of c for a number of possible system parameters such that limiting will not contribute to miss distance.

TABLE 9.5-1 CRITICAL VALUES OF c

	$\mu_h = 0.25$			$\mu_h = 0.5$			$\mu_h = 1.0$		
$\omega_{nh}T$ \ α	2.5	3.5	4.5	2.5	3.5	4.5	2.5	3.5	4.5
10	2.7	2.6	2.5	2.8	2.7	2.6	3.0	2.8	2.7
15	2.5	2.4	2.3	2.5	2.5	2.4	2.7	2.5	2.3
20	2.4	2.3	2.3	2.4	2.4	2.3	2.5	2.4	2.3
30	2.3	2.2	2.2	2.3	2.3	2.2	2.3	2.2	2.2

$\mu_a = 0.5$ and $\omega_{na} = 2\ \omega_{nh}$ in all cases

9.6 MISS DISTANCE DUE TO A TARGET LATERAL ACCELERATION

In this case it has been assumed that the perturbation of the constant bearing trajectory is due to the target executing an evasive manoeuvre equal to a constant g turn T seconds before interception. If the target lateral acceleration is f_t then the component perpendicular to the original sight line is $f_t \cos \phi_o$, and the component of velocity perpendicular to the original sightline is $f_t \cos \phi_o \, t$. When considering heading errors we said that there was an input velocity equal to $U_m \psi_e \cos \psi_{fco}$. Since the system responds to relative motion it matters not whether we regard this disturbance as a missile lateral velocity or target lateral velocity. Hence, the differential equation must be of the same form as equation (9.5-2):

$$(D + \frac{\alpha}{T - t}) z_m = f_t \cos \phi_o \, t \qquad (9.6\text{-}1)$$

The solution to this equation is

$$z_m = \frac{f_t \cos \phi_o \, (T - t)T}{(\alpha - 1)(\alpha - 2)} \{(\alpha - 1)t/T - 1 + (1 - t/T)^{\alpha - 1}\}$$

This can be differentiated twice to yield $\ddot{z}$:

$$\ddot{z}_m = f_t \cos \phi_o \{\frac{\alpha}{\alpha - 2}\} \{1 - (1 - t/T)^{\alpha - 2}\}$$

Hence:

$$f_m = f_t \cdot \frac{\cos \phi_o}{\cos \psi_{fco}} \{\frac{\alpha}{\alpha - 2}\} \{1 - (1 - t/T)^{\alpha - 2}\} \qquad (9.6\text{-}2)$$

Again it is seen that if $\alpha < 2$ then $f_m \to \infty$ as $t \to T$. If the effect of the cosine terms is negligible (e.g. head on or tail chase) it is also seen that if $\alpha = 3$ the initial missile acceleration is zero and the terminal latax is four times the target latax; if $\alpha = 4$ terminal latax is twice that of the target. The limiting case for $\alpha = \infty$ is for the effective missile latax to equal the target latax throughout the engagement.

Figs 9.6-1 to 9.6-3 show missile lateral acceleration as a function of non-dimensional time t/T for $\omega_{nh} T = 20$ and for a biquadratic representation of the system; the full line in each figure shows the response of a lag free system as given in equation (9.6-2). The trajectory therefore is quite different from that due to a heading error. If $\alpha > 2$ then the missile latax decreases as the engagement proceeds for a heading error; in the case of a target lateral acceleration the missile latax increases from zero to a maximum at impact. In the latter case limiting will occur, if it occurs at all, near and at the end of the engagement. If this occurs for a second or

MISSILE LATERAL ACCELERATION/TIME DUE TO A TARGET LATERAL ACCELERATION

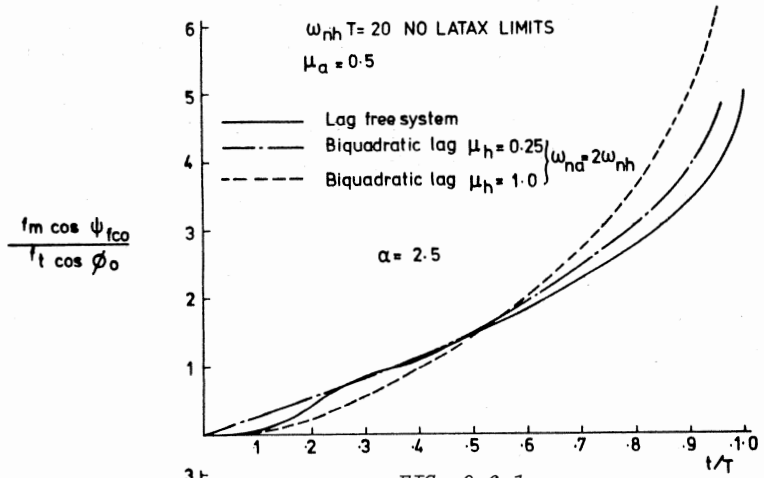


FIG. 9.6-1

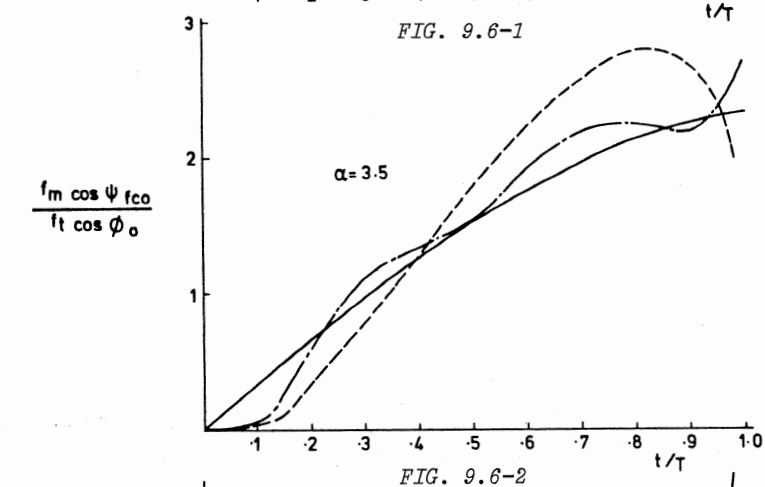


FIG. 9.6-2

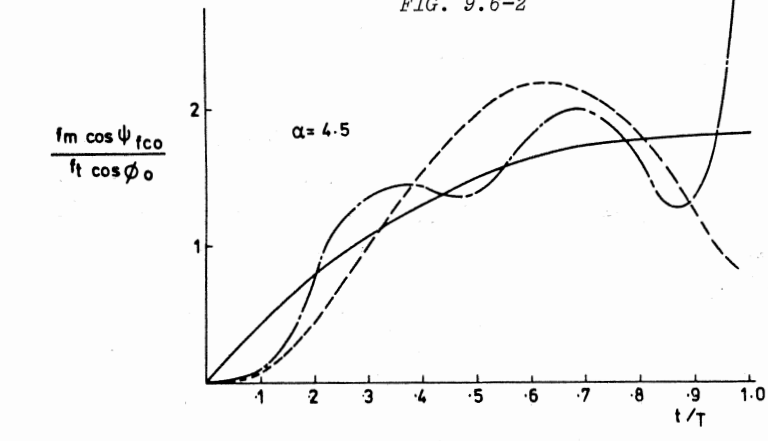


FIG. 9.6-3

so the effect on system accuracy can be disastrous.

The miss distance M_a due to a target lateral acceleration commenced at T seconds before impact is shown in Figs 9.6-4 to 9.6-12. It was thought desirable to put some limits on demanded lateral acceleration. For $\alpha = 2.5$ the lag-free effective terminal missile latax is five times the target effective latax and simulation has shown that in a system with dynamic lags, if the limits are set to 25% in excess of this then limiting does not result in any serious degradation of accuracy. For $\alpha = 3.5$ and $\alpha = 4.5$ the limits have been set to lower values but in both cases 25% in excess of the respective terminal lag-free value. Since setting the limits appreciably below these values would have resulted in large miss distances it was considered better to set them at the minimum acceptable value for accuracy. Setting these limits any higher would have had only a very minor effect on the shape of these graphs.

It is seen from the figures that $\alpha = 2.5$ is not good for intercepting manoeuvring targets, especially if the main lag in the system is well damped. The response is too sluggish. Also, of course, the missile terminal latax is very high. A value of α between 3.5 and 4 appears to be best. Values much above this result in an excessively oscillatory response especially if the main lag is underdamped. There is a nice balance between good damping to reduce oscillations in the response and too much damping which delays the response especially with low values of α, hence favouring the target which manoeuvres a second or so before impact. The optimum time for the target to start a manoeuvre against a homing system whose autopilot is between two and five times as fast as the homing head is T seconds before impact where $\omega_{nh} T \simeq 8$. If $\omega_{nh} = 4$ this is 2 seconds before impact. If the autopilot response is about the same as the homing head the target has more latitude in time to make a successful manoeuvre and the resulting miss distance can be much greater.

The subtle difference between homing guidance using proportional navigation and command guidance should now be apparent. The latter system needs a great deal of sophistication and computing in order to reduce miss distance for fast approaching targets to a metre or so. In the absence of noise homing systems are capable of reducing miss distances to zero because the loop gain approaches a high value towards the end of the trajectory, always assuming that the system has enough time for transients to decay. The great strength of this type of guidance is that missile latax tends to zero at the end of

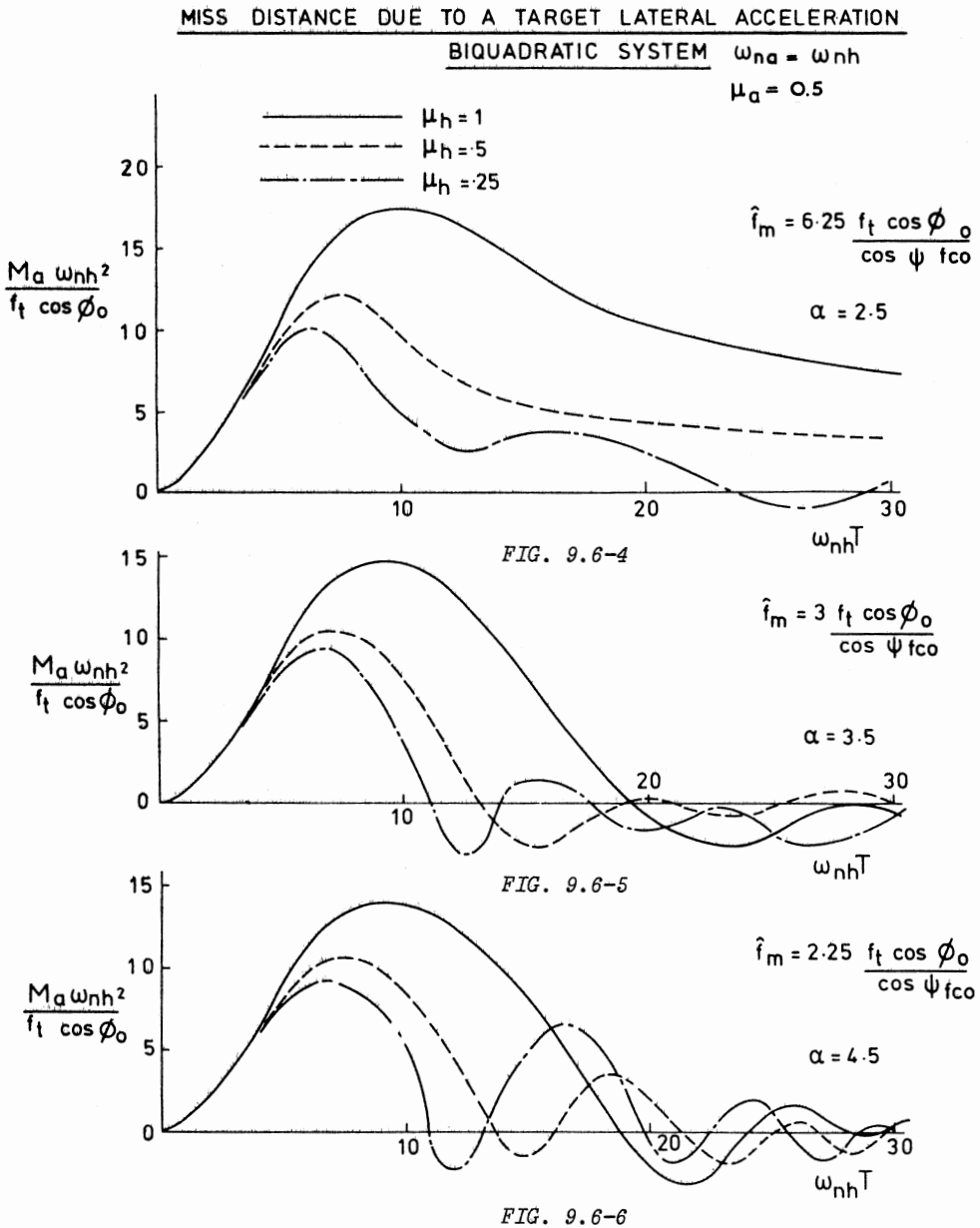


FIG. 9.6-4

FIG. 9.6-5

FIG. 9.6-6

MISS DISTANCE DUE TO A TARGET LATERAL ACCELERATION

BIQUADRATIC SYSTEM $\omega_{na} = 2\omega_{nh}$

$\mu_a = 0.5$

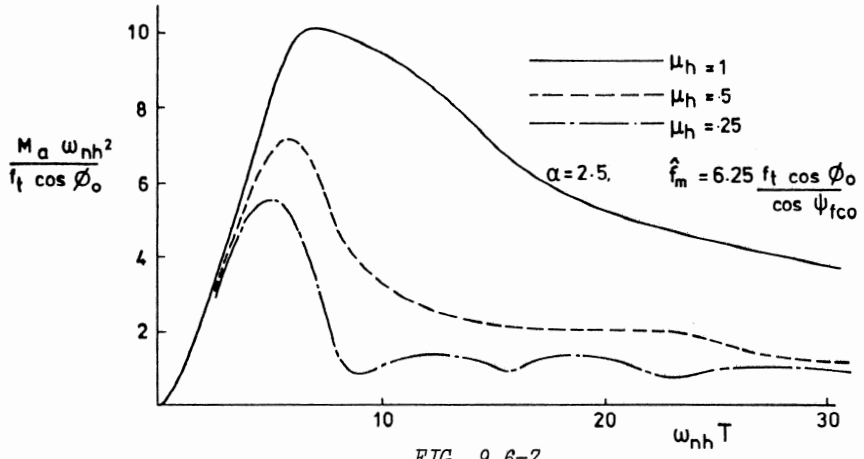


FIG. 9.6-7

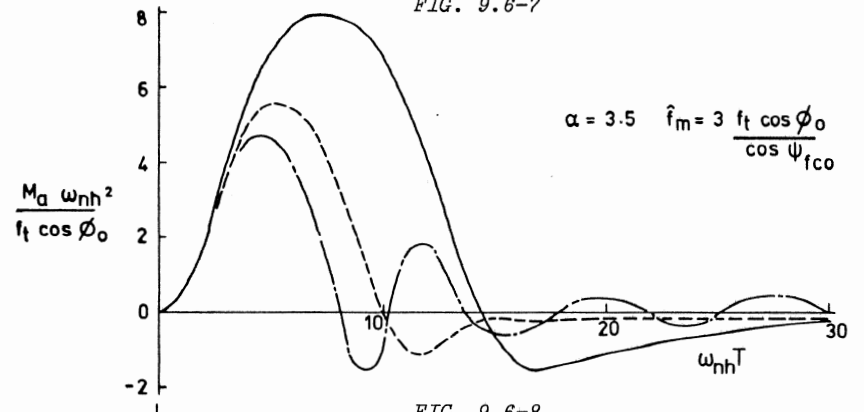


FIG. 9.6-8

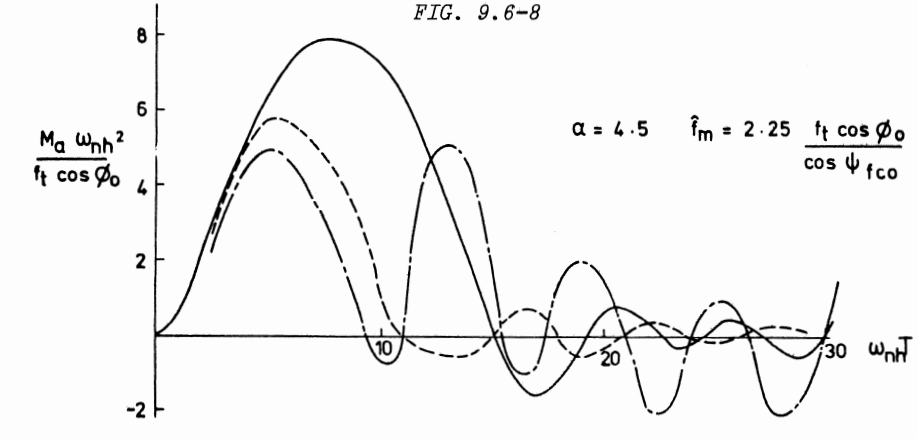


FIG. 9.6-9

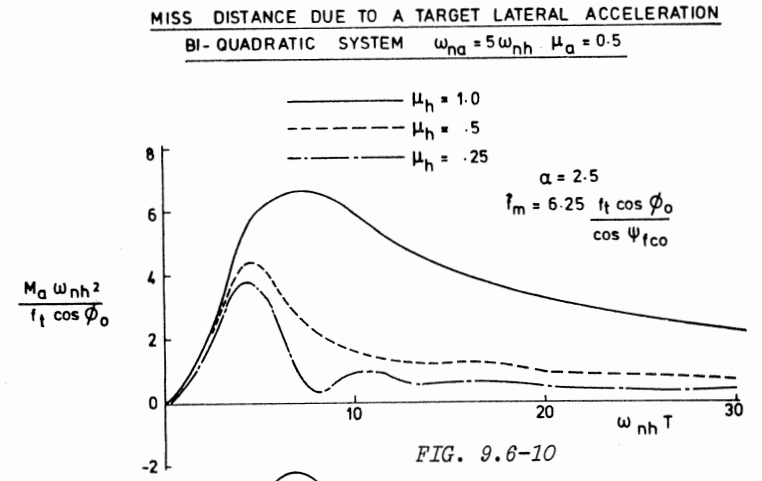


FIG. 9.6-10

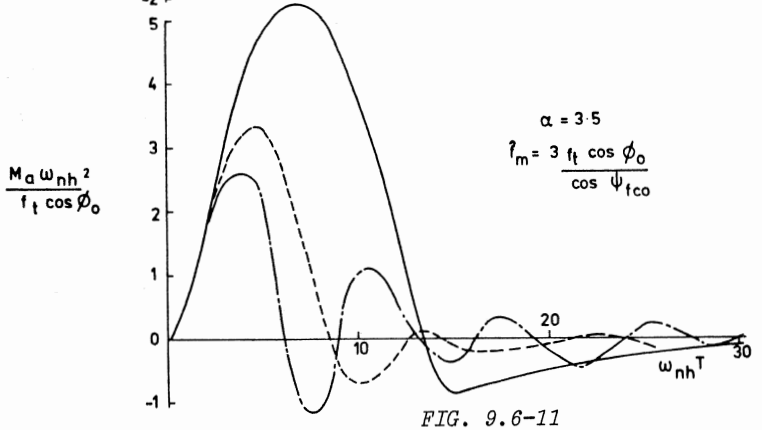


FIG. 9.6-11

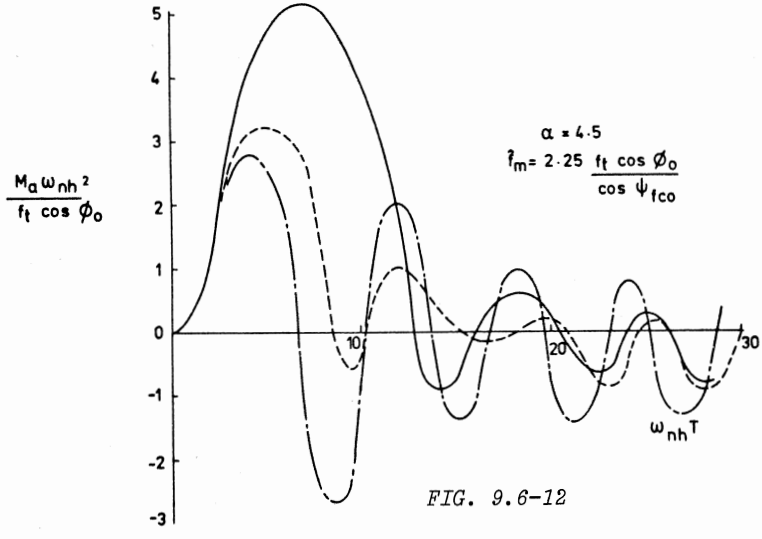


FIG. 9.6-12

the engagement due to a heading error leaving the missile's full g capability to deal with target evasive manoeuvre. The weakness of homing is that even with a well designed system the guidance law is not all that clever at dealing with target evasive manoeuvre. In a conventional command system the effect of a target manoeuvre of ng is likely to increase or decrease the existing missile latax by approximately that amount. Proportional navigation requires the missile to execute more or much more than this towards the end of the engagement.

9.7 MISS DISTANCE DUE TO ANGULAR NOISE

White noise due to a given signal-to-noise ratio and the receiver characteristics in general is assumed to remain sensibly constant a few seconds before impact. Its spectrum is $\Phi_a(\omega) = K_a^2$ rad^2/rad/sec = a constant. Since now the "input" to the system is angular noise, in any analogue simulation it must enter the closed loop at a point which is the analogue of sight line angles. However, if one is using the kinematic gain α then the analogue of sight line angle is not immediately available, see Fig 9.7-1. Since $\tau\dot{r} = r$

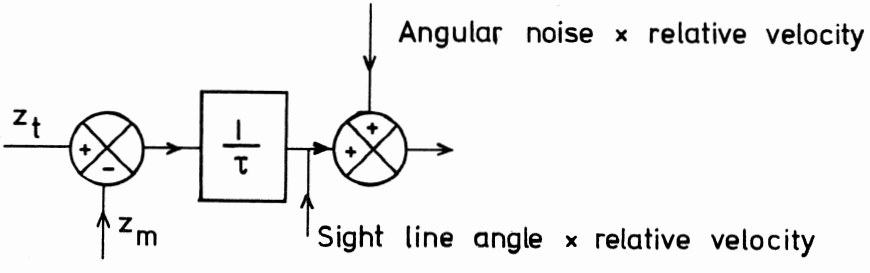


FIG 9.7-1 Angular noise entering a homing system

then the noise "angle" can be inserted as shown provided it is realised that it is scaled accordingly. In other words since we have the analogue of sight line angle times relative velocity it becomes immediately apparent that the virtual input must be proportional to relative velocity.

The control engineer sees this as part of a block diagram where z_m is the output and the feedback transfer function is $1/\tau$ i.e. as $\tau \to 0$ the feedback gain $\to \infty$. This means that provided the system remains stable and the system

speed of response is high the output should → 0 as $\tau \to 0$. Put in another way, when the range to go is large a given angular input can be interpreted as a large apparent linear movement of the target. As the range decreases the apparent linear movement decreases also. Indeed at zero range to go the apparent target and real target will coincide. Hence, since a zero lag system is capable of availing itself of information right up to zero time to go, one would expect an infinitely fast system to exhibit zero miss due to angular noise. In real systems there must be a miss due to angular noise because there are lags in the system and instability must set in just before impact; the miss is usually small since the signal-to-noise ratio just before impact is generally very good in homing systems. Indeed in both active and passive systems the total distance between energy source and receiver is virtually zero at impact. In semi-active systems the signal-to-noise ratio is usually good at impact since for approaching targets the distance from illuminator to target is a minimum and the reflected path length is zero. If a zero lag system exhibits zero miss then one would expect a zero bandwidth system to exhibit infinite miss due to angular noise. Bain (3) comes to this conclusion using a generalised mathematical approach. However one notes that in the general solution he obtains it is implied that the time of engagement is much longer than the effective time constant of the system. This would imply an $(\text{infinity})^2$ time of engagement, a difficult concept! Since at very long times to go the open loop gain is virtually zero it can be visualised that the deviation of a low bandwidth missile from the correct trajectory due to angular noise would tend to increase to a very large value, from which it cannot effectively recover. The results shown in Fig 9.7-2 have been obtained on an analogue computer for 10 second engagements and for homing head bandwidths in the range 2-8 rad/sec. It is seen that within these limits the r.m.s. miss due to angular noise M_n is proportional to the relative velocity, is proportional to the square root of the spectral density of the noise, and is inversely proportional to the bandwidth of the dominant lag. Poor damping, as ever with noise inputs, has a bad effect on system performance; in fact, if the damping ratio of the dominant mode is reduced much below 0.5 the effect tends to be catastropic, and this has already been noted by Bain (3). It will be seen that with a slow autopilot the miss is further increased. This is to be expected since the overall system bandwidth is decreased. However careful inspection of the computer runs reveals that the increase in miss is partly due to the decreased stability of such a system. Results have been obtained for α = 1, 2, 3, 4, 5, and three values of ω_{nh}. One hundred runs were made

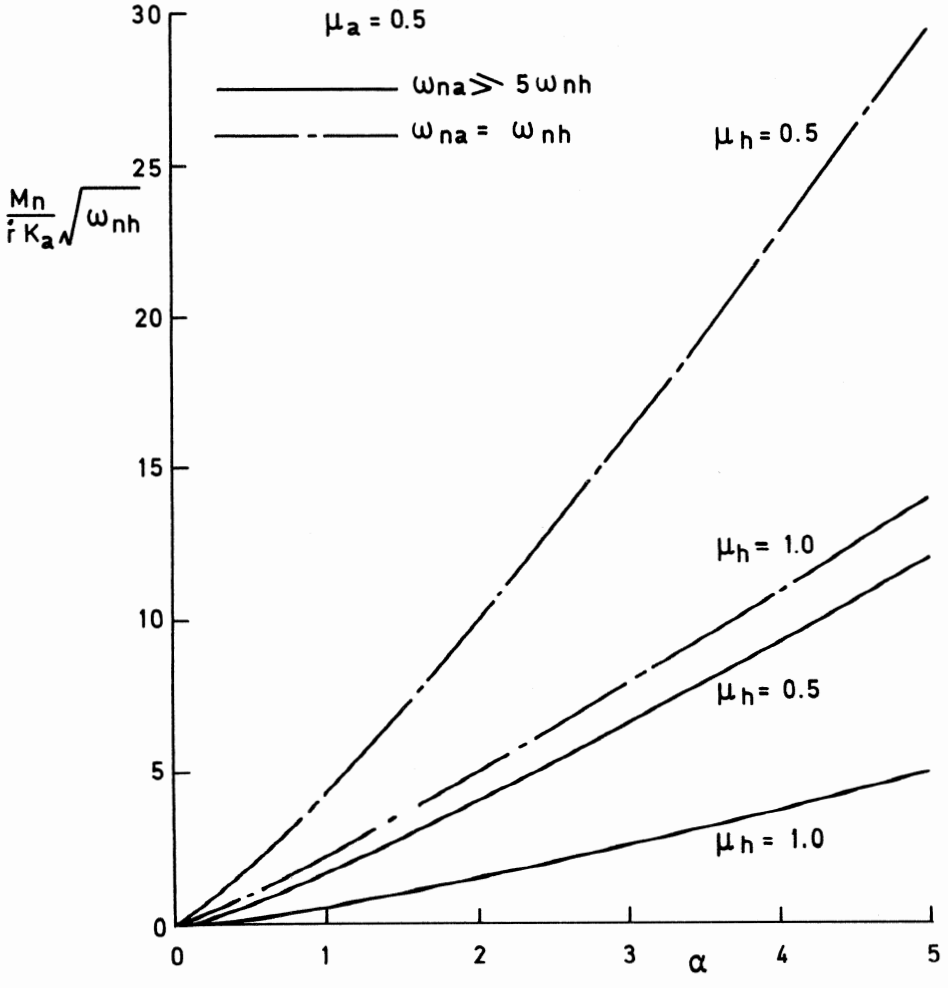


FIG 9.7-2 Miss distance due to angular noise

for each point on a graph in order to obtain a reasonable approximation to the r.m.s. miss, a total of six thousand computer runs in all.

A figure often quoted for K_a^2 in many design studies using radar is 10^{-7} rad^2/rad/sec. If $\dot{r}$ is about M = 2.5 (850 m/sec) then the r.m.s. miss due to angular noise for α = 4 and ω_{nh} = 8 can be in the range 0.4 to 2m depending on damping and autopilot bandwidth.

9.8 MISS DISTANCE DUE TO GLINT

We have already seen that glint is an apparent movement of a target due to motion causing differential phase changes at the receiver. As with command

systems ideally we require the missile to filter out all this spurious information. Unfortunately it is mainly a very low frequency phenomenon whose bandwidth is typically equal to or less than that of our system. This means that in command systems provided the guidance loop is reasonably well damped the r.m.s. miss is not likely to be greater than the r.m.s. value of the glint itself. In homing systems we have seen that no attempt is made to compensate for range and therefore a given apparent linear movement of the target results in ever increasing noise due to glint as range to go decreases. The miss then will tend to increase with the bandwith of the system; the feedforward gain $\to \infty$ as $\tau \to 0$, not the feedback gain. If now the glint spectrum is defined by

$$\Phi_g(\omega) = \frac{K_g^2}{1 + \omega^2 T_g^2}$$

where K_g^2 has the dimensions m^2/rad/sec and T_g is a constant typically about 0.25 secs we can define L_g, the r.m.s. value of the glint as

$$L_g^2 = K_g^2 \int_o^\infty \frac{d\omega}{1 + \omega^2 T_g^2} = \frac{\pi}{2} \frac{K_g^2}{T_g}$$

Since the ratio of the system bandwidth to the noise bandwidth is clearly relevant we can define this as $\omega_n / 1/T_g = \omega_n T_g$; a value of $\omega_n T_g$ in excess of unity means that the system bandwidth is larger than the noise bandwidth, and vice-versa.

With these parameters in mind a number of analogue computer runs have been made and the r.m.s. values for miss distance M_g are shown in Fig 9.8-1. These results clearly show that in homing systems the r.m.s. miss due to glint tends to be greater than the r.m.s. value of the glint itself. A high kinematic gain and low damping will increase the miss. A slow autopilot will also increase the miss despite the reduction in effective bandwidth; again this is almost certainly due to an earlier onset of loop instability. The two sets of graphs also show that if the system bandwidth is less than the noise bandwidth considerable filtering will occur resulting in a smaller miss. Fig 9.8-1(b) could well be typical of a system where $\omega_{nh} = 8$ say and T has been effectively reduced to 1/16 sec due to frequency agility in the receiver. However it must be pointed out that in practice, the behaviour of a real system homing on to a real glinting target is often more complicated than depicted above. If the glint is severe, limiting may well occur somewhere in the system a second or so before impact; this will tend to reduce the miss but could increase the miss if the missile is already trying to pull

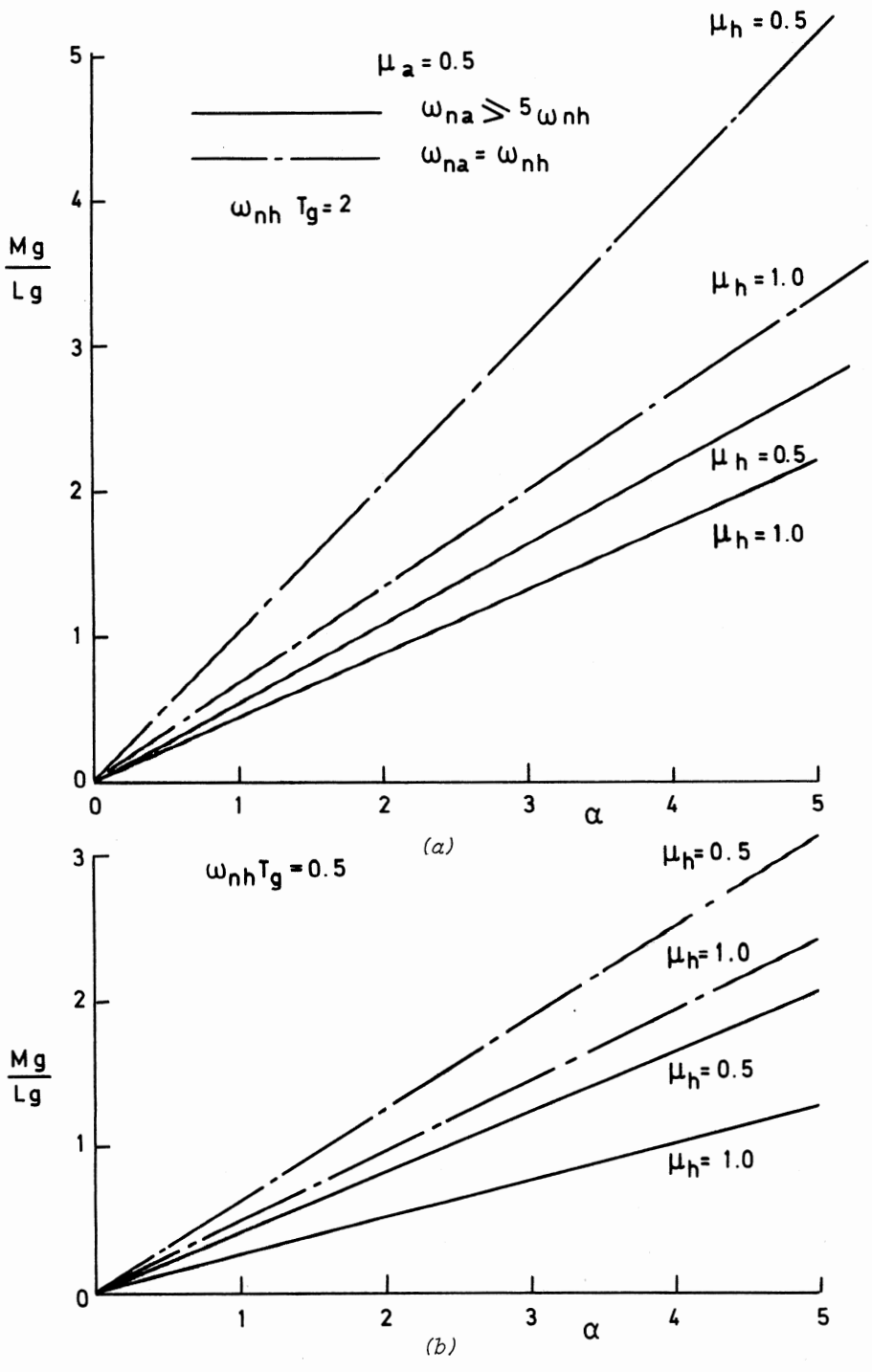


FIG 9.8-1 Miss distance due to glint

maximum g in order to catch a manoeuvring target. Also in some simulations when a real homing head has been used plus an illuminator and a physical model of the target, an increase in the bandwidth of the system has actually resulted in a reduction in the miss due to glint; and the characteristics of the automatic gain control have also been found to be relevant. Apparently a higher bandwidth system spends more of its homing time in areas of high signal strength and this affects accuracy.

Nevertheless, the above arguments must indicate general trends. What is not in doubt is that in systems using radar, glint is often the main contributor to miss distance in homing systems and that the use of frequency agility can significantly reduce its effect.

9.9 ACCELERATION VECTORED NAVIGATION

Possibly the most severe kinematic condition that can be presented to a missile guidance-control system is an air-to-air engagement with a crossing target. The range could be short or very short, such that the missile could still be boosting at zero range to go. There may be no time to aim ahead; indeed one may wish to avoid this complication anyway. There are two certain ways of failing to hit a target (a) to be unable to pull sufficient g a second or so before impact and (b) to exceed the maximum mechanical angle of look the homing head can achieve at any time. Also since the missile speed can be varying considerably throughout the engagement and one is possibly thinking of all-round attack if one assumes a value for the relative velocity it is likely to be very approximate. Acceleration vectored navigation is a method which offers some improvements on conventional proportional navigation under such circumstances.

In the conventional system we demand a lateral acceleration $f_y = K\,\dot{\theta}$ either assuming or measuring the relative velocity, see equation 9.3-7. Ideally we want the acceleration vector normal to the sight line *MT*, see Fig 9.9-1. We have already noted that any component of acceleration (target or missile) directed along *MT* will merely hasten or retard impact. The missile could easily be boosting at 20g. If the angle of look is 30^{o} then there is a component normal to the sight line of 10g. This is equivalent to a lateral manoeuvre and will occur in the absence of any sight line spin; it is therefore a system bias and since there are only two points relevant to the guidance picture it is equivalent to a 10g target manoeuvre perpendicular to the sight line. If the kinematic gain is 4 then this will require the missile to pull

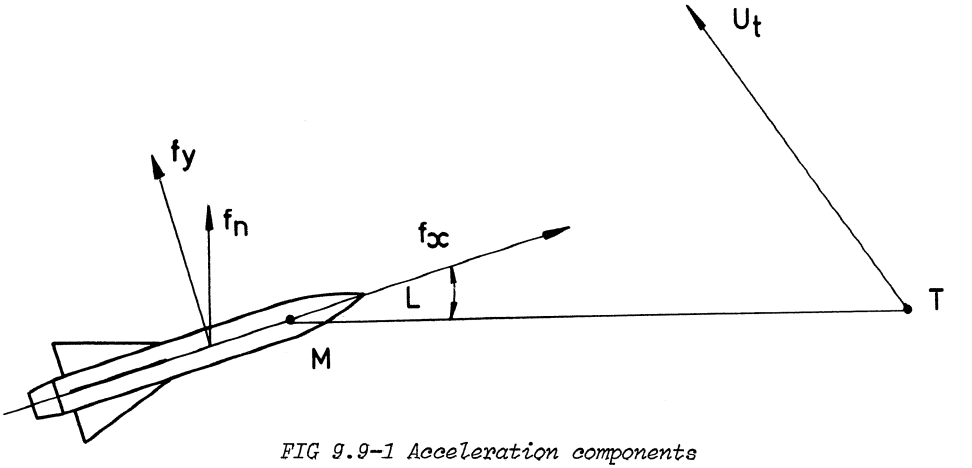


FIG 9.9-1 Acceleration components

20g at the end of the engagement just to overcome this bias. Hence in the guidance law we could include a term to account for the forward acceleration f_x and the look angle L. Since

$$f_n = f_x \sin L + f_y \cos L$$

and we demand $f_y = K\dot{\theta}$ then

$$f_n \simeq f_x L + K\dot{\theta} \text{ for small values of } L$$

For zero bias another term in the guidance equation is required thus

$$f_y = K\dot{\theta} + K_1 L \text{ where } K_1 = -f_x$$

$$\text{or } f_y + f_x L = K\dot{\theta} \qquad (9.9\text{-}1)$$

This is the same as saying that a lateral acceleration and a look angle must be demanded at the same time. This can be achieved if we feed back both these quantities in the correct proportions. Angle of look L can be measured by a potentiometer between the head and the body. Fig 9.9-2 shows a simplified

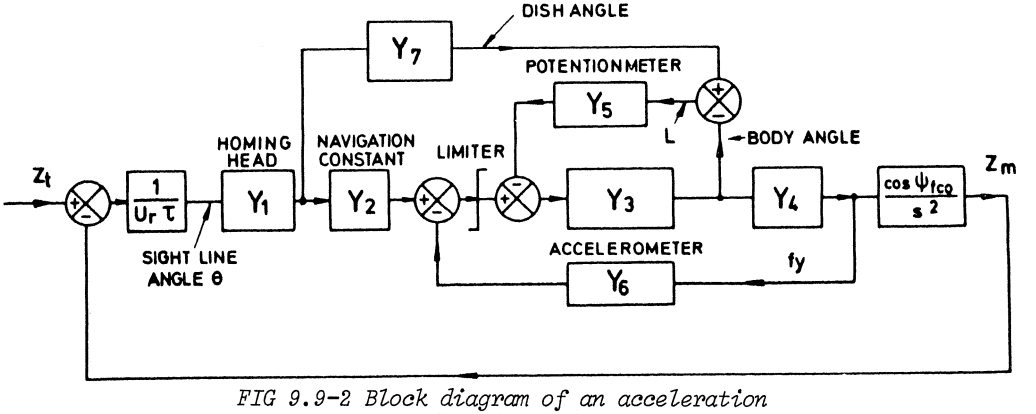


FIG 9.9-2 Block diagram of an acceleration vectored navigation system

block diagram of such a system where

Y_1 is the homing head transfer function, sight line to sight line rate.

Y_2 is the navigation constant.

Y_3 is the transfer function including the servo and aerodynamics to produce body angle.

Y_4 is the transfer function from body angle to lateral acceleration.

Y_5 is the potentiometer gain which must vary with f_x.

Y_6 is the accelerometer transfer function.

Y_7 is the transfer function, sight line rate to dish direction.

Since the loop containing Y_3 and Y_5 is a look angle demand loop, the input to this can be limited so that the angle of look is not exceeded. The engineering of such a system involves much technical know-how. However, such a system has been successfully developed using thrust vector control in order to achieve maximum manoeuvrability at launch. In addition it has been found to give a better coverage than a conventional proportional navigation system, and is less sensitive to the real value of relative velocity.

9.10 AN INTEGRATED FORM OF PROPORTIONAL NAVIGATION

We have seen that it is essential to establish an accurate measure of sight line rate uncontaminated by missile body motion; any such coupling reduces stability margins and degrades the accuracy of the homing process. Ideally we need autopilot instruments to regularise the missile response and rate gyros on the head to assist in decoupling the head from missile motion. The integrated form of proportional navigation seeks to reduce the number and cost of missile-borne instruments and uses a guidance law of the form

$$\psi_f = k\,\theta \qquad (9.10\text{-}1)$$

instead of the conventional form

$$\dot{\psi}_f = k\,\dot{\theta}, \text{ see equation (9.3-10)}$$

Provided the initial conditions are the same, the two types of proportional navigation should be indistinguishable.

Fig 9.10-1 shows the angles which are relevant. The angle channel receiver produces a signal proportional to $(\theta - d)$. If a potentiometer is mounted between the head and the body we have a measure of $(d - \psi_m)$. If a free gyro is mounted in the missile we have a measure of ψ_m with respect to the original launch direction. To establish θ we can write

$$(\theta - d) + (d - \psi_m) + \psi_m = \theta \qquad (9.10\text{-}2)$$

This means that in order to obtain a signal proportional to sight line angle

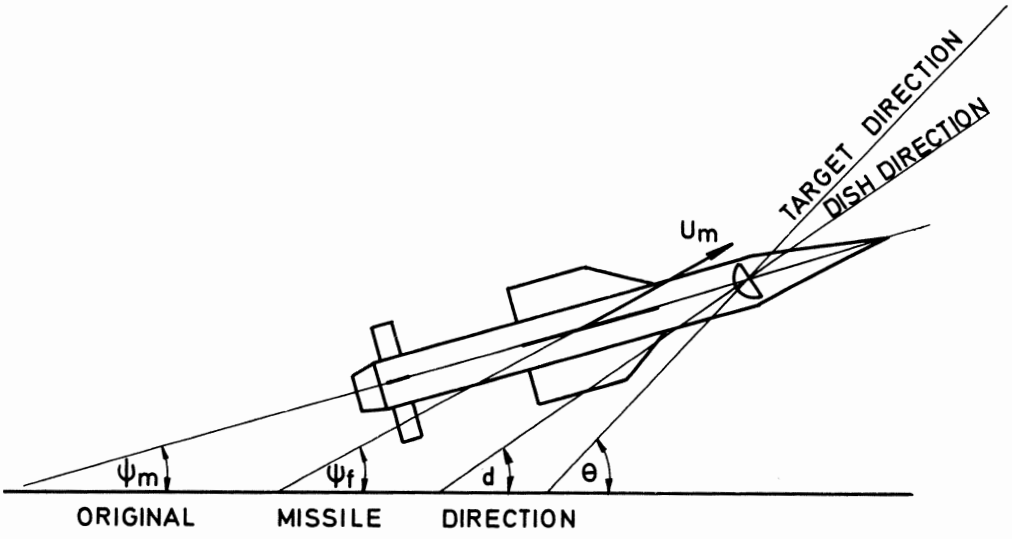


FIG 9.10-1 Definition of angles

with respect to a space datum we need to feed back positively a signal proportional to $(d - \psi_m)$ and another proportional to ψ_m. The simplified block diagram is shown in Fig 9.10-2 for small perturbations about a constant

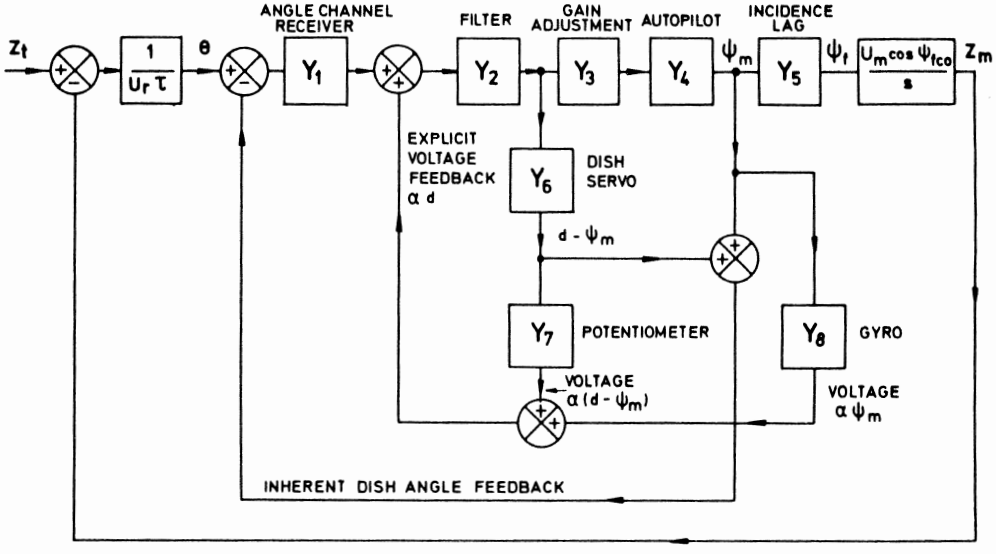


FIG 9.10-2 Block diagram of an integrated form of proportional navigation

bearing course. The signal from the receiver is passed through a filter to reduce the high frequency noise content. This signal is then fed to a servo which drives the dish relative to the missile; it is also used as a demand to the autopilot for a change in body direction. The autopilot therefore uses body angle feedback provided by a free gyro. The complete autopilot including fin servos, aerodynamics gyro and any compensation used is shown as a block with a transfer function Y_4. The gyro signal is used again to provide a signal proportional to dish angle when summed with the potentiometer signal, so although the gyro is used in the autopilot block it is shown separately with a transfer function Y_8. The reader should have no difficulty in verifying that there is no feedback of dish angle or body angle provided $Y_1 = Y_7 = Y_8$. Hence the system is virtually open loop from θ to ψ_f (except of course for the gyro feedback in the autopilot) and therefore we can write:

$$\frac{\psi_f}{\theta} = Y_1 \cdot Y_2 \cdot Y_3 \cdot Y_4 \cdot Y_5 = Y = \frac{\dot{\psi}_f}{\dot{\theta}}$$

The navigation constant can be adjusted by adjusting Y_3. If the d.c. gain of Y is k then

$$\alpha = \frac{k\, U_m \cos \psi_{fco}}{U_r}, \text{ see equation (9.3-10)}$$

The following points should be noted:

(a) The complete guidance loop contains one integration instead of two integrations and one inherent differentiation - no real change.

(b) The homing head lag does not figure as one of the system lags; instead there is the incidence lag, the autopilot lag as before, and any filtering one may design into Y_2 and Y_3.

(c) The difficulties in matching Y_1, Y_7 and Y_8 should not be underestimated. Any lack of linearity or bias will lead either to some effective negative feedback which could reduce the open loop gain significantly, or conversely could lead to some nett positive feedback which could lead to instability.

(d) Any drift in the gyro should have a negligible effect. 0.1^{o}/sec associated with a missile speed of 573 m/sec is an effective latax of only 1 m/sec and a reasonable quality gyro should drift much less than this.

An obvious choice of this type of proportional navigation would be for a sea skimming missile using a radio altimeter for height control and a good quality free gyro for heading control. If the end-course guidance is homing, then the same gyro can then be used for both the pitch and yaw guidance channels.

REFERENCES

1. Cornford E.C. and Bain R.W. The kinematics of proportional navigation courses for a missile with a time lag. Royal Aircraft Establishment Tech Note No G.W. 85 October 1950.

2. Bain R.W. and Trebble F.E. Proportional navigation for a missile with a quadratic time lag. Royal Aircraft Establishment Tech Note No G.W. 307 April 1954.

3. Bain R.W. The analysis of linear homing navigation systems. Royal Aircraft Establishment Tech Note No G.W. 427 August 1956.

4. Jenkins D.P. Proportional navigation with a quadratic lag missile. Royal Radar Establishment Memorandum No 1250 October 1957.

5. Jenkins D.P. Proportional navigation against a weaving target. Royal Radar Establishment Memorandum No 1572 May 1959.

6. Heap E. Methodology of research into command-line-of-sight and homing guidance. AGARD Lecture Series No 52 on Guidance and control of tactical missiles May 1972.

CHAPTER 10
WIENER FILTER THEORY APPLIED TO GUIDANCE LOOP DESIGN

10.1 INTRODUCTION

To date almost all guided weapon control and guidance systems have been designed using what is now termed "classical" control theory. This is that body of knowledge dating from the work done by Maxwell on governors in 1868, and continuing through to the 1930s when the contributions from the Bell Telephone Laboratories in the USA, and in particular those of Nyquist and Bode, brought the theorists and practitioners together with such success. Due to the need for the highly accurate positioning of radars and guns during World War II, these classical techniques were applied by engineers and scientists from many disciplines and by the end of the war they had been developed to the point where no further significant progress was possible without a complete change of approach to control system design.

It is to the credit of these classical techniques that they have continued to serve the control engineer admirably up to the present time and there is every indication that they will continue to be a basic part of every control engineer's tool kit for many years to come.

The essential characteristic of control system design using classical methods is the ad-hoc approach, where each effect, e.g. stability, noise or non-linearity, is usually considered separately and in turn and a circular design procedure is continued until a suitable compromise has been reached.

Consider the design of a missile guidance loop. Although the objective is to minimise the miss distance at interception this is seldom stated explicitly in the design procedure. Instead a particular guidance loop configuration is chosen, and a sufficiently large loop gain is specified in order to satisfy steady state accuracy requirements. The loop is then stabilised using compensating networks until the Nyquist diagram (or its equivalent) indicates that acceptable gain and phase margins have been achieved.

The effect of noise on system performance is then studied, either by calculation or simulation, and if this is found to result in an unacceptable miss or possible saturation in some component then extra filtering must be introduced and/or the loop gain reduced. Since this will probably result in an increase in miss distance adjustments must again be made in this area, the most likely result being a compromise between noise performance, stability and

steady state accuracy. Other effects would be considered as necessary. For example, if large signals are likely to make the use of linear system approximations invalid then these effects would also have to be studied, either by some technique such as the describing function or again by simulation.
That this approach has worked, and worked well, for over three decades places classical control theory in a strong position and the need for an alternative approach is by no means obvious. However, the clue to this need can be found in the expressions we have used previously, "acceptable gain and phase margins", "compromise between noise performance and stability". What are the most acceptable gain and phase margins and what is the best compromise? Even when the system design has been finalised and no amount of parameter adjustment seems to result in a better system can we be sure that the best system really has been produced - in our case the one that minimises the miss distance?
In some fields a few per cent difference between an optimum and sub-optimum system would not matter but for a guided missile such a reduction in miss distance might result in a considerably greater percentage improvement in the weapon system lethality.
"Modern" control theory, as opposed to classical control theory, is concerned with the explicit optimisation of some parameter, or parameters, e.g. the miss distance, and it attempts to achieve this end without going through the ad-hoc procedure necessitated by the use of classical techniques. Of course, this need to attain, or at least define, the optimum was of as much concern to engineers 30 years ago as it is today and indeed modern control theory had its origins in World War II with the development of the Wiener filter. Although its practical use is somewhat limited in the GW field, the Wiener filter is developed here not just because of its historical importance but because it introduces many of the ideas of modern control and provides a valuable link with the modern concepts of state estimation and optimal control.

10.2 THE WIENER FILTER

A fundamental problem facing engineers in the 1940s was how to design a radar that would track a target in the presence of noise, e.g. glint or thermal noise, with the least possible error. Their first problem was to define what they meant by least possible error. One interpretation might be the least possible value of:

$$\int_0^T e \, dt$$

where e is the instantaneous error and tracking takes place in the interval $0 < t < T$.

This, however, would not be a satisfactory measure if e had both positive and negative excursions since the integral could then be small whilst e could have a large rms value. This problem could be overcome by integrating the absolute value of the error but in fact the measure of "performance index" (PI) chosen at the time was:

$$\int_0^T e^2 \, dt$$

This was justified on the grounds that large errors would be penalised more than small ones but in truth the chief reason for this choice was that the PI was mathematically tractable and it is still very popular today for the same reason.

The next problem was to define in some way the likely behaviour of the target and the noise. Obviously, if either the target motion or noise is a well defined function of time then there is no problem. Normally, however, they will be random functions of time as illustrated in Figure 10.2-1.

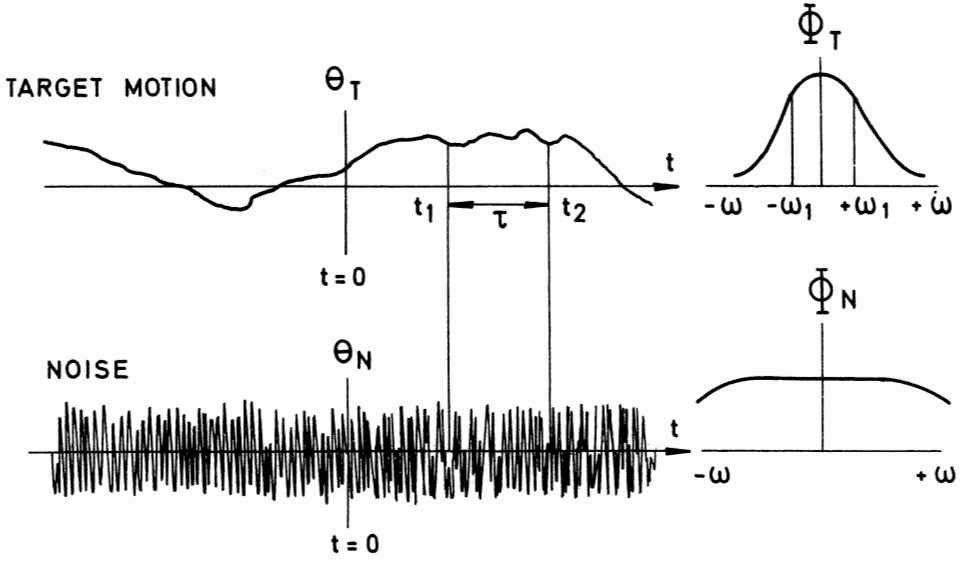


FIG 10.2-1 Target motion and noise characteristics

Both are clearly random and yet each has characteristics which enable it to be distinguished from the other. The target motion signal obviously has most of its power concentrated in the lower frequencies (an aircraft, say, cannot move quickly from one position to another because of its large inertia). We can also see that because of this the value of θ_T at some time t_1 is to a certain extent correlated with the value of θ_T at time t_2. As $\tau = (t_2 - t_1)$ increases this correlation becomes less but nevertheless it still exists. On the other hand the noise θ_N is assumed to have its power distributed equally over a wide band of frequencies and there is very little correlation between the value of θ_N at time t_1 and the value at time t_2. Only as τ becomes very small does any correlation begin to exhibit itself.

Now provided these two signals, θ_T and θ_N, are stationary, which means that their statistical characteristics do not vary with time, then all the information we need to design a Wiener filter is incorporated in the power spectra (shown to the right in Figure 10.2-1), or the equivalent autocorrelation functions, of the signal and noise. It should be noted that the part of the power spectrum corresponding to negative frequencies is just as "real" as that for positive frequencies. Consider the generation of a sine wave produced by the horizontal projection of a rotating vector, positive frequencies are produced by anticlockwise rotation and negative frequencies by clockwise rotation. The total power due to a frequency component ω_1 is thus the sum of the powers attributed to $+\omega_1$ and $-\omega_1$.

In the original filter conception, Wiener made use of the autocorrelation functions of θ_T and θ_N and he also accounted for any cross-correlation between θ_T and θ_N, as might be the case, for instance, if θ_N is glint noise generated by target motion. This approach, however, requires much tedious mathematics and although completely general it does tend to mask the physical operation of the filter. We shall assume no correlation between θ_T and θ_N and shall use only the power spectra representations of the two signals, following the simplified derivation of the Wiener filter given by Bode and Shannon (1). Since we are primarily concerned with the design of GW systems let us assume that the required output of our filter is θ_M, the angular position of the missile we wish to guide (Figure 10.2-2).

The input to the filter is $\theta_T + \theta_N$ and we will assume that we have available, from previous tests and trials, a knowledge of Φ_T and Φ_N, the power spectra of the target motion and noise. If we further assume that we are dealing with a CLOS or beam riding system then our objective is to find the filter F,

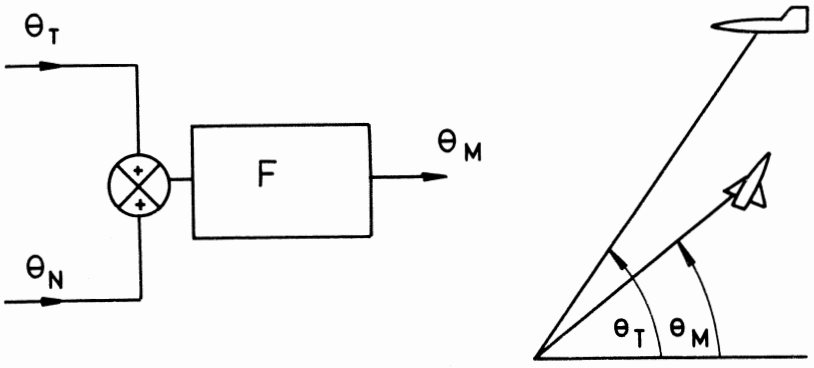


FIG 10.2-2 Filter F

which represents the complete closed loop system relating θ_M to the input $\theta_T + \theta_N$, that will minimise the PI.

$$\int_0^T (\theta_M - \theta_T)^2 \, dt$$

where T is the time of flight. That is, we want a system that will keep the missile on the target sight line with the minimum squared error for the whole time of flight. Just from inspection of the power spectra Φ_T and Φ_N it is clear that some form of low pass filter must result. Instead of trying various low pass filters of different bandwidths until the best one has apparently been obtained our objective will be to concentrate on the minimisation of the PI and trust that the best filter will emerge from the mathematics - this is the essence of the modern approach.

10.3 WIENER FILTER DERIVATION

In order to make use of the given information, viz the power spectra of θ_T and θ_N, the PI must be changed to a form capable of representing signals and noise in the frequency domain. To do this we must consider a PI of the form

$$\overline{e^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} (\theta_M - \theta_T)^2 \, dt$$

i.e. we must consider θ_M and θ_T for all time, both past and future. Only in this way can we make use of the power spectra information which cannot be used to represent signals like that shown in Figure 10.3-1 which is clearly non-stationary. It is zero, or at least undefined, outside the interval $-T < t < T$. Although necessary in order to make any progress, the assumption that we have knowledge of θ_M and θ_T for all time will clearly have implications with regard to the practical realisation of any filter that is derived. We

FIG 10.3-1 A non-stationary signal

will pursue this point later.

If we now make one final assumption, that F is to be linear, then the alternative form of the PI can be written

$$\overline{e^2} = \int_{-\infty}^{\infty} \{|F|^2 \Phi_N + |1 - F|^2 \Phi_T\} \, d\omega$$

where the arguments in ω have been omitted for brevity.

The first term of the integrand represents the contribution to the error power due to noise. The second term of the integrand represents the contribution to the error power due to the fact that F is not, in general, unity and therefore, even with no noise present, θ_M will not follow θ_T exactly.

Let $F = Ae^{j\phi}$, A and ϕ being the gain and phase shift of the filter as functions of frequency,

therefore $\overline{e^2} = \int_{-\infty}^{\infty} \{|Ae^{j\phi}|^2 \Phi_N + |1 - Ae^{j\phi}|^2 \Phi_T\} d\omega$

$$= \int_{-\infty}^{\infty} \{(A^2\cos^2\phi + A^2\sin^2\phi)\Phi_N + (\left[1 - A\cos\phi\right]^2 + A^2\sin^2\phi)\Phi_T\} d\omega$$

$$= \int_{-\infty}^{\infty} \{A^2\Phi_N + A^2\Phi_T - 2A\Phi_T \cos\phi + \Phi_T\} d\omega$$

This will be a minimum if $\cos\phi$ is a maximum i.e. $\phi = 0$. Also, adding and subtracting $\frac{\Phi_T^2}{\Phi_N + \Phi_T}$ to the integrand we get:

$$\overline{e^2} = \int_{-\infty}^{\infty} \{A^2 \left[\Phi_N + \Phi_T\right] - 2A\,\Phi_T + \frac{\Phi_T^2}{\Phi_N + \Phi_T} + \Phi_T - \frac{\Phi_T^2}{\Phi_N + \Phi_T}\} d\omega$$

$$= \int_{-\infty}^{\infty} \{\left[A\sqrt{\Phi_N + \Phi_T} - \frac{\Phi_T}{\sqrt{\Phi_N + \Phi_T}}\right]^2 + \frac{\Phi_T \Phi_N}{\Phi_N + \Phi_T}\} d\omega$$

This will be a minimum if

$$A\sqrt{\Phi_N + \Phi_T} - \frac{\Phi_T}{\sqrt{\Phi_N + \Phi_T}} = 0$$

$$F = Ae^{j\phi} = \frac{\Phi_T}{\Phi_N + \Phi_T} \qquad (10.3\text{-}1)$$

is the optimum filter that will minimise the chosen PI.

Unfortunately this result has been derived without regard to physical realisability since, in the derivation, we assumed a knowledge of θ_T and θ_N for all time ($-\infty < t < \infty$).

The implication of this can be seen more clearly if it is remembered that the inverse Fourier transform of equation (10.3-1) is the impulse response of the required filter in the time domain.

This impulse response will have both positive and negative tails (Figure 10.3-2) and the negative part cannot be utilised unless it is known before hand what value the input to the filter is going to be in the future so that the contribution from the negative part of its impulse response can be computed *now*.

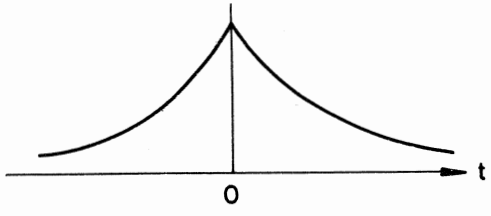


FIG 10.3-2 Impulse response of unrealisable filter

Thus the optimal filter given by equation (10.3-1) can only be implemented in practice provided the whole time history of $\theta_T + \theta_N$ is available. This implies waiting until the end of the engagement before deciding how to control the missile (θ_M) from launch.

This is illustrated in Figure 10.3-3 where two impulse components (A and B) of the input signal $\theta_T + \theta_N$ and their respective impulse responses are shown. Thus the output of the filter θ_M, at some arbitrary time $t = 0$, is made up not only of the positive tail of the impulse response due to A but also the negative tail of the impulse response due to B. Unfortunately, however, if we want to deduce θ_M *now* ($t = 0$) we must forgo the contribution due to the negative tail of B unless, of course, we can wait α seconds which in our particular application isn't usually possible. It is equivalent to saying that we want the line of sight θ_M to be where the line of sight θ_T was α seconds ago.

It would seem then that the way out of this difficulty would be just to consider the positive tail of the impulse response. This would certainly be

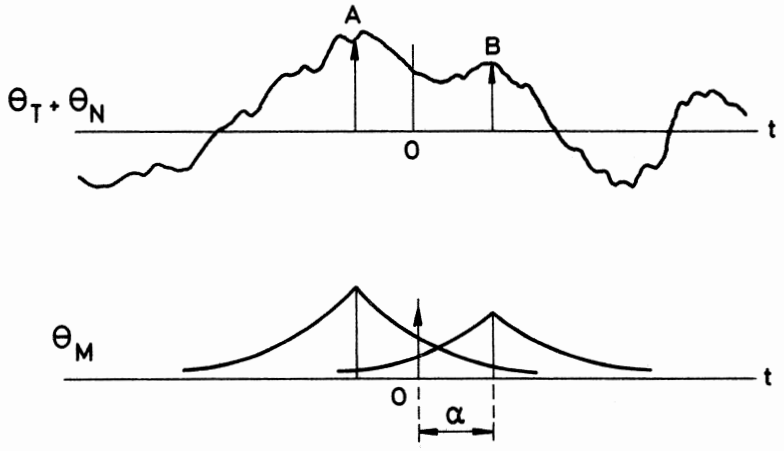


FIG 10.3-3 Impulse responses due to components A and B of filter input

physically realisable since it would not rely on knowing values of $\theta_T + \theta_N$ that had yet to occur. It would not, however, minimise $\overline{e^2}$.

The crux of Wiener filter design is to recognise that although we might not know the value of the impulse *B* exactly (without waiting those α seconds), because the signal $\theta_T + \theta_N$ has some correlation with the signal α seconds hence it should be possible to deduce what the contribution from the negative tail of the impulse response due to *B* α seconds into the future is most likely to be now. Wiener's achievement was to show how to do this.

Following the Bode-Shannon derivation, we will start with the optimum but not physically realisable filter given by equation (10.3-1) and observe that the denominator can be factored into two parts, $(\Phi_N + \Phi_T)^+$ which has all its poles and zeroes in the upper half plane and $(\Phi_N + \Phi_T)^-$ which has all its poles and zeroes in the lower half plane. We can then consider our filter *F* split into two cascaded parts, F_1 and F_2 (Figure 10.3-4).

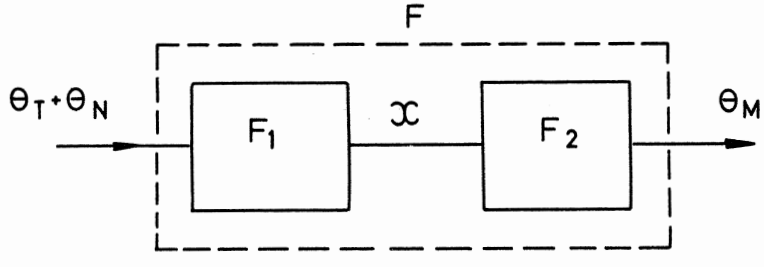


FIG 10.3-4 Filter represented by two cascaded parts

If we make $F_1 = \dfrac{1}{(\Phi_N + \Phi_T)^+}$ then the power spectrum of the output x of F_1 will be:

$$\Phi_x = |F_1|^2 \ (\Phi_N + \Phi_T) = 1$$

If we now paid no regard to physical realisability we would have to make

$$F_2 = \frac{\Phi_T}{(\Phi_N + \Phi_T)^-}$$

in order that $F = F_1.F_2$ as given by equation (10.3-1). However, since the input to F_2 is now white noise (Φ_x = a constant) and is therefore uncorrelated from one instant to the next the value of x at any time in the future is uncorrelated with the value *now* and we are thus justified in ignoring the negative tail of the inverse transform of F_2 since the contribution to θ_M *now* due to future values of x are equally likely to be positive or negative, the best estimate of the sum of these contributions due to future values of x being zero.

The physically realisable form of equation (10.3-1) is therefore given by

$$F_{pr} = \frac{1}{(\Phi_N + \Phi_T)^+} \left[\frac{\Phi_T}{(\Phi_N + \Phi_T)^-} \right]_+ \qquad (10.3\text{-}2)$$

where $[\ ]_+$ indicates that only that part of F_2 having poles in the upper half plane (and therefore having an inverse transform equal to zero for $t < 0$) is considered.

Example 10.3-1

To find the physically realisable filter that minimises $\overline{(\theta_M - \theta_T)^2}$ where θ_T is contaminated with noise θ_N and the respective power spectra of θ_T and θ_N are $\Phi_T = \dfrac{3}{1 + \tau^2\omega^2}$ and $\Phi_N = 1$.

$$\Phi_N + \Phi_T = 1 + \frac{3}{1 + \tau^2\omega^2} = \frac{4 + \tau^2\omega^2}{1 + \tau^2\omega^2} = \frac{(2 - \tau j\omega)(2 + \tau j\omega)}{(1 - \tau j\omega)(1 + \tau j\omega)}$$

$$(\Phi_N + \Phi_T)^+ = \frac{2 + \tau j\omega}{1 + \tau j\omega} \text{ and } (\Phi_N + \Phi_T)^- = \frac{2 - \tau j\omega}{1 - \tau j\omega}$$

$$F_1 = \frac{1}{(\Phi_N + \Phi_T)^+} = \frac{1 + \tau j\omega}{2 + \tau j\omega}$$

$$F_2 = \left[\frac{\Phi_T}{(\Phi_N + \Phi_T)^-} \right]_+$$

$$= \left[\frac{3}{(1 + \tau j\omega)(2 - \tau j\omega)} \right]_+$$

$$= \left[\frac{1}{1 + \tau j\omega} + \frac{1}{2 - \tau j\omega}\right]_+$$

$$= \frac{1}{1 + \tau j\omega}$$

Hence, the physically realisable filter is given by

$$F_{pr} = F_1 \cdot F_2 = \frac{1}{2 + \tau j\omega}$$

or, writing s for $j\omega$

$$F_{pr} = \frac{1}{2 + \tau s} \qquad (10.3\text{-}3)$$

It is interesting to note that if just the positive tail of the inverse transform of F given by equation (10.3-1) had been taken we would have obtained:

$$\left[F\right]_+ = \left[\frac{\Phi_T}{\Phi_N + \Phi_T}\right]_+ = \left[\frac{3}{4 + \tau^2\omega^2}\right]_+ = \frac{\frac{3}{4}}{2 + \tau j\omega}$$

$$\text{or } \left[F\right]_+ = \frac{\frac{3}{4}}{2 + \tau s}$$

$\left[F\right]_+$ is thus not the same as the physically realisable optimum. The increased gain of F_{pr} compared to $\left[F\right]_+$ $(1{:}\frac{3}{4})$ can be interpreted physically as accounting for the most likely extra contribution that the negative tails of the impulse responses due to future values of $\theta_T + \theta_N$ would make to θ_M now.

10.4 THE CONSTRAINED WIENER FILTER

In order to highlight the essential features of the Wiener filter it has been necessary to simplify the design problem as much as possible. In particular, the PI would probably not be a practical one since we have chosen to minimise the mean square error without regard to any constraints that might have to be imposed on the system in practice. For instance, it would be necessary to ensure that we are not demanding unacceptable levels of lateral acceleration, say, in order to achieve this minimum.

Any extra system restrictions on some variable θ (velocity, acceleration etc) can be incorporated in the PI by letting it take the form:

$$PI = \int_0^T \{(\theta_M - \theta_T)^2 + \lambda\, \theta^2\}\, dt$$

Although this only provides a soft limit on θ, sensible choice of the weighting factor λ will ensure that the desired limits on θ are exceeded for only a small percentage of the time. Also, the power spectrum representation of signals will only be valid for a sufficiently long time of flight T, a

necessary condition for all applications of the Wiener filter.
Let us assume that θ is related to θ_M by a transfer function $1/H$ (Figure 10.4-1).

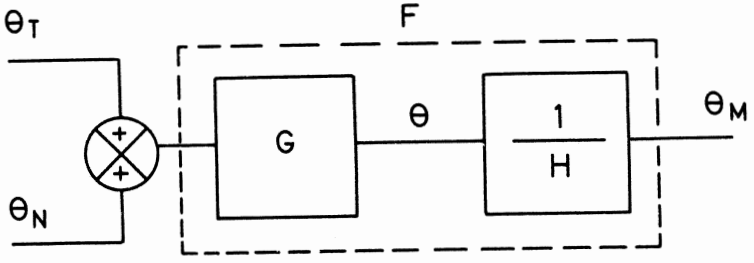


FIG 10.4-1 Variable θ to be constrained

The required optimum filter F has thus been split into two parts, the output θ of the first part being the variable we wish to constrain. Following the same approach as for the unconstrained Wiener filter we note that the PI can be expressed in its equivalent frequency domain form, the total mean error power being given by:

$$\overline{e^2} = \int_{-\infty}^{\infty} \{\left|\frac{G}{H}\right|^2 \Phi_N + |1 - \frac{G}{H}|^2 \Phi_T + \lambda|G|^2 (\Phi_N + \Phi_T)\}d\omega$$

By expanding and completing the square as done previously it can be shown that

$$F = \frac{G}{H} = \frac{\Phi_T}{(1 + \lambda |H|^2)(\Phi_N + \Phi_T)}$$

and the physically realisable optimum is given by

$$F_{pr} = \frac{1}{(1 + \lambda |H|^2)^+ (\Phi_N + \Phi_T)^+} \cdot \left[\frac{\Phi_T}{(1 + \lambda |H|^2)^- (\Phi_N + \Phi_T)^-}\right]_+$$

Example 10.4-1

Using the same data as for the unconstrained Wiener filter but now assuming that $\dot{\theta}_M$ is to be limited to some extent by minimising the PI:

$$\int_o^T \{(\theta_M - \theta_T)^2 + \lambda \dot{\theta}_M^2\}dt$$

Hence $H = S$

and $1 + \lambda|H|^2 = 1 + \lambda (j\omega)(-j\omega) = (1 - \sqrt{\lambda}j\omega)(1 + \sqrt{\lambda}j\omega)$

$(1 + \lambda |H|^2)^+ = (1 + \sqrt{\lambda}j\omega)$; $(1 + \lambda|H|^2)^- = (1 - \sqrt{\lambda}j\omega)$

$$F_{pr} = \frac{(1 + \tau j\omega)}{(1 + \sqrt{\lambda} j\omega)(2 + \tau j\omega)} \left[\frac{3}{(1 + \tau j\omega)(1 - \sqrt{\lambda} j\omega)(2 - \tau j\omega)} \right]_+$$

$$= \frac{(1 + \tau j\omega)}{(1 + \sqrt{\lambda} j\omega)(2 + \tau j\omega)} \left[\frac{A}{(1 + \tau j\omega)} + \ldots\ldots \right]_+$$

$$\text{where } A = \frac{3}{(1 - \sqrt{\lambda}\left[-\frac{1}{\tau}\right])(2 - \tau\left[-\frac{1}{\tau}\right])} = \frac{\tau}{\tau + \sqrt{\lambda}}$$

$$F_{pr} = \frac{\frac{\tau}{\tau + \sqrt{\lambda}}}{(1 + \sqrt{\lambda} s)(2 + \tau s)} \qquad (10.4\text{-}1)$$

by writing s for $j\omega$.

It is seen that this differs from the unconstrained filter (10.3-3) by the addition of the factor

$$\frac{\frac{\tau}{\tau + \sqrt{\lambda}}}{(1 + \sqrt{\lambda} s)}$$

The significance of this term will become apparent in the next chapter although at the moment we can note that it has arisen as a result of our desire to constrain the control variable (if we consider our control of $\dot{\theta}_M$ to be the way we influence the motion of the missile) whereas the factor $\frac{1}{(2 + \tau s)}$ is concerned with the filtering of the input $\theta_T + \theta_N$.

Although the example chosen does correspond to a practical missile system, e.g. the velocity controlled anti-tank missile, it is far more likely that any fully automatic CLOS system would use acceleration control ($H = s^2$). Furthermore, the spectra for the target motion and noise signals, even assuming that they can be considered to be stationary random signals, will almost certainly be more complex than those used in the illustrative example. Burt and Bain [2] produced the definitive treatment of the application of Wiener filter theory to GW systems and their work has recently been extended to include coloured (bandlimited) noise (3).

However, although the method of solution for these more realistic Wiener filter formulations is, in principle, no different from that used for our simple example, the filter complexity rapidly becomes overwhelming. Furthermore, the filter obtained represents the overall closed loop response. The guidance loop filter itself must still be extracted from this and the end result is a filter which the control engineer finds difficult to interpret and, if necessary, modify.

Recent developments in modern control theory have led to the recognition that the filtering (or estimation) and controlling aspects which are so inextricably

mixed together in the Wiener filter can be separated out. This means that although the objective is still the minimisation of a PI the engineer can now identify definite functions with the parameters in the final design.

REFERENCES

1. BODE H.W. and C.E. SHANNON A simplified derivation of linear least square smoothing and prediction theory. Proc IRE April 1950.

2. BURT E.G.C. and R.W. BAIN On the statistical optimisation of guided-weapon systems. HMSO 1966.

3. RICHARDS L.J. and PAMELA THORN Optimum Wiener filters for line-of-sight guided weapon systems. RAE Tech Report 72116 1972.

CHAPTER 11

MODERN CONTROL THEORY APPLIED TO GUIDANCE LOOP DESIGN

11.1 INTRODUCTION TO MODERN CONTROL THEORY

In the last chapter we saw that the essence of the modern approach to control system design is to concentrate on the minimisation of a performance index (PI). However, although the objective of the Wiener filter is to minimise a PI it is still not truly representative of the modern techniques that have been developed in the last two decades.

A distinguishing feature of modern control theory is that it is concerned with the representation and manipulation of signals in the *time domain* whereas classical control theory is essentially frequency domain orientated. This reliance by classical theory on the description of signals by their individual frequency components has usually meant that the restrictions of linearity and time invariance must be placed on the systems considered in order that the superposition principle can be utilised, albeit implicitly. By remaining in the time domain we are free from these restrictions but it would seem that we have then reverted to the state of control theory prior to the work of Bode and Nyquist when system design relied on the explicit solution of the differential equations describing the system. In fact, this is so, but now we can enlist the services of the digital computer to solve the equations numerically, a facility not available before the mid 1950s. Indeed, it is true to say that the enormous growth in control theory in recent years has parallelled the growth in the utilisation of the digital computer.

Another important concept in modern control theory is the idea of *system state*. Consider a simplified system such as a point mass missile (Figure 11.1-1).

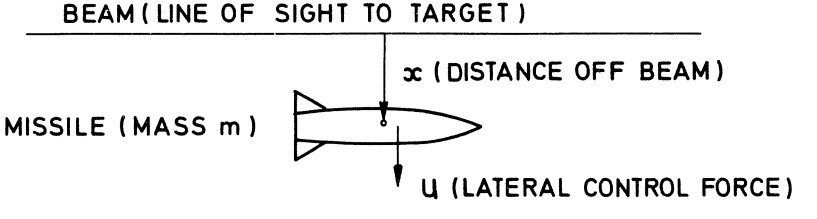


FIG 11.1-1 A simplified missile system

Assume the application of a lateral control force u, resulting for example from a fin movement. The differential equation describing the motion of the missile is therefore

$$u = m\ddot{x} \qquad (11.1\text{-}1)$$

where m is the mass of the missile. The "state" of this second order system is completely determined if the position (x) and velocity ($\dot{x}$) are known at every instant. Accordingly the variables x and $\dot{x}$ are known as the system *state variables*.

Therefore, a two-dimensional vector having components x, $\dot{x}$ defines the state of the system and is known as the system *state vector* and is written:

$$\underline{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$$

Note this use of the word vector. We are really referring to a point in a two-dimensional "state space" and the state of time t_1, say, can be thought of as the tip of a vector emanating from the origin, or state of system equilibrium $x = 0$, $\dot{x} = 0$ (Figure 11.1-2).

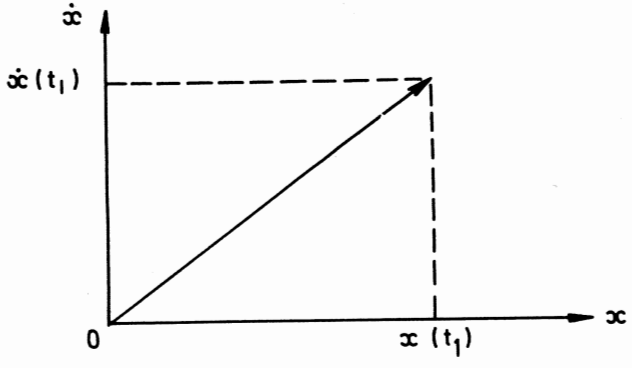


FIG 11.1-2 A state vector in 2-D state space

For a third order system a three-dimensional state space would be required and for an nth order system a n-dimensional state space.

It is important not to be overawed by the mathematical jargon. When we talk of a four-dimensional state space we are just referring to all the possible sets of values that the four states of a 4th order system can take up, these states representing real physical quantities such as position and velocity.

We will now make use of state variables in reformulating equation (11.1-1).

Let $x = x_1$ and $\dot{x} = x_2$

therefore $\dot{x} = \dot{x}_1 = x_2$ (11.1-2)

and $\ddot{x} = \dot{x}_2 = \frac{u}{m}$ (11.1-3)

i.e. we have decomposed the one second order differential equation into two first order differential equations. In this form they are ideally suited to numerical solution by computer. Another reason for considering all the system states, x_1 *and* x_2, and not just one output, say x_1, as would normally be done using classical methods, is concerned with the implementation of optimal control which will be treated shortly.

An alternative way of writing equations (11.1-2) and (11.1-3) would be in vector-matrix form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u$$

The compactness and utility of this notation becomes apparent when high order systems are considered, although it should be noted that it can only be applied to linear systems, i.e. when $\dot{x}_1$ and $\dot{x}_2$ are formed from linear combinations of x_1 and x_2 and have no terms of the form $x_1 x_2$ or x_1^2. Nevertheless, it should be appreciated that non-linear first order differential equations, although not expressible in this compact vector-matrix form, can be handled by the computer with the same facility as linear equations.

Generalising from our second order example, any nth order linear system can be represented by a vector-matrix equation of the form:

$$\dot{\underline{x}} = A\,\underline{x} + B\,\underline{u} \qquad (11.1\text{-}4)$$

where, in general, A and B can have time varying elements, $\underline{u}$ can represent multiple inputs or controls, and any of the n components of $\underline{x}$ can be regarded as system outputs.

This highlights another essential difference between classical and modern control theory. Multi-dimensional inputs or outputs are accommodated quite naturally using the state space approach whereas classical theory is restricted to single-input single-output situations.

11.2 DETERMINISTIC OPTIMAL CONTROL

Our ultimate aim is to solve, using modern control theory, the problem successfully tackled by Wiener filter techniques in *Example 10.4-1*, viz the minimisation of the PI

$$\int_0^T \{(\theta_M - \theta_T)^2 + \lambda\, \dot{\theta}_M^2\}\, dt$$

where θ_T is a random function, the measurement of which is contaminated with noise θ_N.

However, in order to introduce the concept of optimal control we will initially assume that θ_T is fixed and that noise is absent. This is the problem of the deterministic regulator, the "set point" being θ_T which, for convenience, can be taken as zero.

We therefore have the one state system (Figure 11.2-1) defined by:

$$\dot{\theta}_M = u \tag{11.2-1}$$

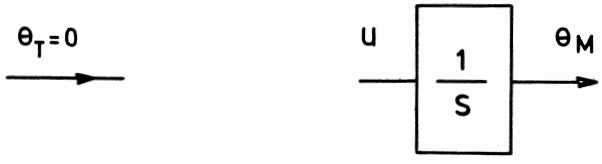


FIG 11.2-1 The one state regulator system

Our objective is to minimise the PI

$$J = \int_{t_o}^{T} \{\theta_M^2 + \lambda u^2\}\, dt \tag{11.2-2}$$

where the lower limit of integration has been changed to t_o for clarity of notation in subsequent equations. However, in order to prepare for the solution of a more general optimal control problem we will consider, initially, a PI of the form:

$$J = P(T)\, \theta_M^2(T) + \int_{t_o}^{T} \{\theta_M^2 + \lambda u^2\}\, dt \tag{11.2-3}$$

The extra term $P(T)\, \theta_M^2(T)$ puts emphasis on the minimisation of the miss distance at the specific time T and would obviously be an appropriate addition to the PI when the interception time T is known in advance (which is generally *not* the case with a CLOS system). Nevertheless, following a heuristic treatment given by Pitman (1), this form of PI does suggest a way of tackling the optimal control problem. In particular, the additional term $P(T)\, \theta_M^2(T)$ is reminiscent of the final value of an integral, viz:

$$\int_{t_o}^{T} \{\frac{d}{dt} P(t)\, \theta_M^2(t)\}\, dt$$

Hence, consider

$$\begin{aligned} \frac{d}{dt} P\theta_M^2 &= \dot{P}\theta_M^2 + 2P\theta_M\dot{\theta}_M \\ &= \dot{P}\theta_M^2 + 2P\theta_M u \end{aligned}$$

where the argument in t has been omitted for brevity.

Therefore

$$\int_{t_o}^{T} \{\frac{d}{dt} P\theta_M^2\}\, dt = \int_{t_o}^{T} \{\dot{P}\theta_M^2 + 2P\theta_M u\}\, dt$$

$$= \int_{t_o}^{T} \{\dot{P}\theta_M^2 + 2P\theta_M u + \theta_M^2 + \lambda u^2\}\, dt - \int_{t_o}^{T} \{\theta_M^2 + \lambda u^2\}\, dt$$

(by adding and then subtracting the term $\theta_M^2 + \lambda u^2$)

$$= P(T)\ \theta_M^2\ (T) - P(t_o)\ \theta_M^2\ (t_o) \qquad (11.2\text{-}4)$$

Hence

$J = P(T)\ \theta_M^2(T) + \int_{t_o}^{T} \{\theta_M^2 + \lambda u^2\}\, dt$ re-stating equation (11.2-3)

$= P(t_o)\ \theta_M^2(t_o) + \int_{t_o}^{T} \{\dot{P}\theta_M^2 + 2P\theta_M u + \theta_M^2 + \lambda u^2\}\, dt$ from equation (11.2-4)

Now in our original PI, equation (11.2-2), we had no term $P(T)\theta_M^2(T)$, i.e. $P(T) = 0$, and hence

$$J = \int_{t_o}^{T} \{\theta_M^2 + \lambda u^2\}\, dt$$

$$= P(t_o)\theta_M^2(t_o) + \int_{t_o}^{T} \{\dot{P}\theta_M^2 + 2P\theta_M u + \theta_M^2 + \lambda u^2\}\, dt \qquad (11.2\text{-}5)$$

Minimising both sides of equation (11.2-5) with respect to u gives:

$$J_{opt} = P(t_o)\theta_M^2(t_o) + \min_{wrt\ u} \int_{t_o}^{T} \{\dot{P}\theta_M^2 + 2P\theta_M u + \theta_M^2 + \lambda u^2\}\, dt$$

since $P(t_o)\theta_M^2(t_o)$ is dependent only on the starting time t_o and is unaffected by u.

Considering the integrand $\{\dot{P}\theta_M^2 + 2P\theta_M u + \theta_M^2 + \lambda u^2\}$, this can be written as a perfect square

$$(K\theta_M + \sqrt{\lambda}\ u)^2 = (\dot{P} + 1)\theta_M^2 + 2P\theta_M u + \lambda u^2$$

provided $K^2 = \dot{P} + 1$ (11.2-6)

and $2K\sqrt{\lambda} = 2P$ (11.2-7)

If equations (11.2-6) and (11.2-7) are satisfied then the integrand is always non-negative and the minimum possible value (i.e. zero) of the integral term will be achieved provided

$$u = u_{opt} = -\frac{1}{\sqrt{\lambda}} K\ \theta_M = -\frac{P}{\lambda}\ \theta_M \qquad (11.2\text{-}8)$$

i.e. u_{opt} is the control that makes the integrand zero for $t_o < t < T$.

Hence u_{opt} can be found provided P can be determined. Eliminating K from equations (11.2-6) and (11.2-7) results in a first order differential equation in P given by

$$\dot{P} = \frac{P^2 - \lambda}{\lambda} \text{ where } P(T) = 0$$

Although this particular non-linear differential equation can be solved analytically, in general it would be necessary to obtain a numerical solution and this is the approach followed here. Starting from the known boundary condition $P(T) = 0$, the differential equation is integrated backwards in time to t_o using a digital computer and the variation of $\frac{P}{\lambda}$ with time is shown in Figure 11.2-2. This figure also illustrates the essential feedback nature of the optimal control given by equation (11.2-8).

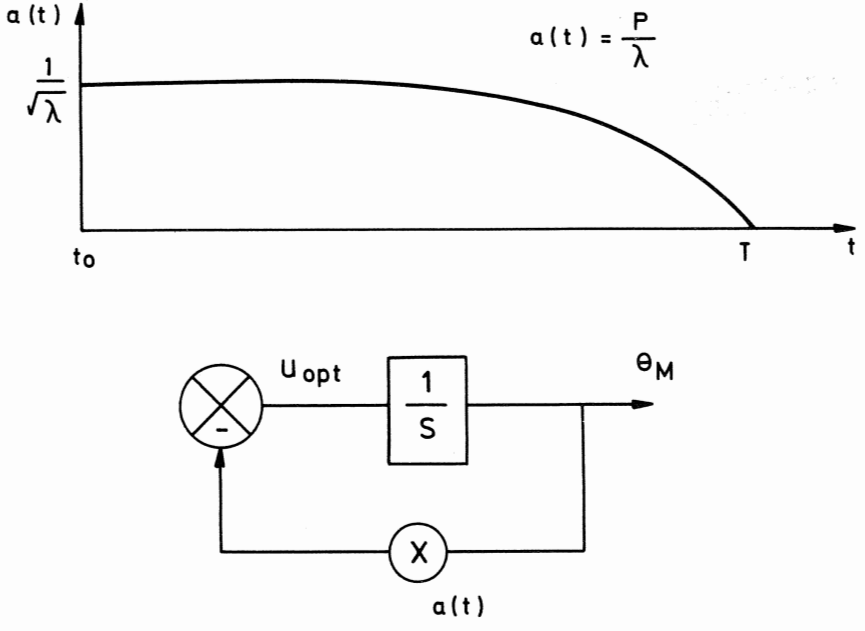


FIG 11.2-2 Time varying optimal feedback gain

It is interesting to observe that as $(T - t) \to \infty$, $u_{opt} \to -\frac{1}{\sqrt{\lambda}}\theta_M$. In other words, if there is a long time to go before interception, at $t = T$, the feedback gain can be kept constant at $\frac{1}{\sqrt{\lambda}}$. Only just prior to interception need the feedback gain be reduced. The significance of this reduction in gain is that with a short time to go little change can be made to θ_M, no matter what

value of u is chosen, because of the lag associated with the system integration (Figure 11.2-2), whereas a reduction in u will at least reduce its own contribution to the PI.

This reduction in feedback gain just prior to interception is contrary to established control system design. It has resulted from our choice of PI which made no allowance for minimising the final value of the state, $\theta_M(T)$, but required only that minimisation be carried out over the interval t_o to T. For CLOS and beam riding missiles this would be a reasonably satisfactory PI and, indeed, it is unlikely that *a priori* knowledge of T would even be available in most cases.

The relative merits of time varying and constant ($\dot{P} = 0$) feedback gains in a CLOS system for PIs with and without weighting on the final states have been studied by Gardner (2). In most cases Gardner found little difference in performance between systems using optimal, sub-optimal or constant gains. Nevertheless, in situations where the intercept time T is known and where a specific weighting on the final states is included in the PI then optimal control systems can provide significant theoretical improvements over systems designed by classical methods. This does not mean that such systems could, or should, necessarily be implemented in practice. However, optimal control theory does provide the engineer with greater insight into the design problem and often suggests gain variations (probably sub-optimal) which would not otherwise have been apparent. An example of this is treated in 11.4 where the optimal control of a homing missile is considered.

In conclusion, it should be noted that although this section has dealt with a specific and simple example, the derivation of the required optimal controller did highlight three essential features which are common to all linear multivariable optimal control problems with quadratic performance indices, viz:

(a) The optimal control is a linear combination of all the states of the system. In classical control parlance, it is a feedback (and sometimes a feedforward) control. (cf $u_{opt} = -\frac{P}{\lambda}\theta_M$)

(b) The optimised PI is dependent only on the initial states. In fact, it is possible to generalise further and say that the optimised PI has a quadratic form (cf $J_{opt} = P(t_o)\theta_M^2(t_o)$)

(c) The derivation of the optimal controller requires the solution of a non-linear differential equation. (cf $\dot{P} = \frac{P^2 - \lambda}{\lambda}$)

The derivation of the optimal control for the general linear multivariable

system with quadratic performance index can be found in appendix A, the treatment being a heuristic one based on the preceding example.

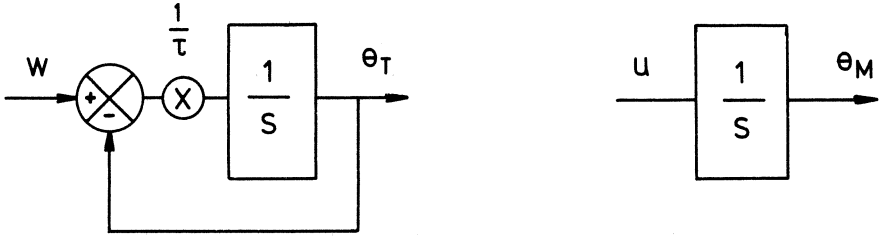
11.3 STOCHASTIC OPTIMAL CONTROL

We will now consider the problem of minimising the PI

$$\int_{t_o}^{T} \{(\theta_M - \theta_T)^2 + \lambda \dot{\theta}_M^2\}\, dt$$

where θ_T is no longer fixed but is a zero mean random function with power spectrum Φ_T. We thus have a stochastic control problem since we are attempting to construct a controller that will minimise a PI when the future behaviour of the system is subject to some uncertainty, in this case the effect of the random input θ_T. This type of problem, where the PI itself is also subject to uncertainty by virtue of the random variables in the integrand, must be tackled by minimising the *expected* value of the PI. Fortunately, provided the uncertainties in the system can be represented by zero mean white noise sources, "shaped" if necessary by appropriate filters, then the design proceeds as for the deterministic optimal controller and the results of appendix A can be used directly. The justification for this can be found in many recent textbooks, e.g. reference 3.

Once again it will be assumed that θ_T can be measured exactly, i.e. $\theta_N = 0$. The case when θ_T is contaminated with noise is deferred until the next chapter. We therefore have the two state system (Figure 11.3-1).



PART OF SYSTEM REPRESENTING TARGET MOTION

PART OF SYSTEM REPRESENTING MISSILE TO BE CONTROLLED.

FIG 11.3-1 The two state stochastic system

where W represents zero mean white noise of power per unit bandwidth q^2 such that

$$\Phi_T = \frac{q^2}{(1 + \omega^2\tau^2)}$$

$$\text{let } \theta_T = x_1 \text{ and } \theta_M = x_2$$

$$\text{therefore } \dot{x}_1 = -\frac{1}{\tau} x_1 + \frac{1}{\tau} W$$

$$\dot{x}_2 = u$$

Hence, re-stating the problem in vector-matrix form, we must find a control u that will minimise the PI.

$$J = \int_o^T \{ \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \lambda u^2 \} \, dt \qquad (11.3\text{-}1)$$

(cf equation (A2), appendix A)

for the system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\tau} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} \frac{1}{\tau} \\ 0 \end{bmatrix} W \qquad (11.3\text{-}2)$$

(cf equation (A1), appendix A)

Now provided W is zero mean white noise (with, strictly, a gaussian amplitude distribution) then the last term of equation (11.3-2) can be ignored as far as the minimisation of the expected value of J is concerned and the procedure developed in appendix A can be applied directly.

Hence

$$\underline{u}_{opt} = -R^{-1} B^T P \underline{x}$$

where $P \equiv \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$ is given by the solution of the Riccati equation

$$-\dot{P} = Q + PA + A^T P - PBR^{-1} B^T P$$

Since P is symmetric ($p_{21} = p_{12}$) this matrix equation defines *three* independent simultaneous differential equations in the independent variables p_{11}, p_{21} and p_{22} which can be obtained by substituting for Q, R, A and B from equations (11.3-1) and (11.3-2) thus giving:

$$\begin{bmatrix} -\dot{p}_{11} & -\dot{p}_{21} \\ -\dot{p}_{21} & -\dot{p}_{22} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} -\frac{p_{11}}{\tau} & 0 \\ -\frac{p_{21}}{\tau} & 0 \end{bmatrix} + \begin{bmatrix} -\frac{p_{11}}{\tau} - \frac{p_{21}}{\tau} & 0 \\ 0 & 0 \end{bmatrix} - \frac{1}{\lambda} \begin{bmatrix} p_{21}^2 & p_{21}p_{22} \\ p_{21}p_{22} & p_{22}^2 \end{bmatrix}$$

Therefore

$$-\dot{p}_{11} = 1 - \frac{2p_{11}}{\tau} - \frac{p_{21}^2}{\lambda} \qquad (11.3\text{-}3a)$$

$$-\dot{p}_{21} = -1 - \frac{p_{21}}{\tau} - \frac{p_{21}p_{22}}{\lambda} \qquad (11.3\text{-}3b)$$

$$-\dot{p}_{22} = 1 - \frac{p_{22}^2}{\lambda} \qquad (11.3\text{-}3c)$$

Since there is no weighting on the value of the states at interception ($t = T$) included in the PI then $P(T) = 0$, i.e. equations (11.3-3) must satisfy the boundary conditions $p_{11}(T) = 0$, $p_{21}(T) = 0$ and $p_{22}(T) = 0$. Hence (11.3-3) can be solved numerically by integrating backwards from the known boundary conditions (Figure 11.3-2). Not too surprisingly, $\dot{P} \to 0$ as the time to go before interception increases thus implying constant control gains for most of the flight.

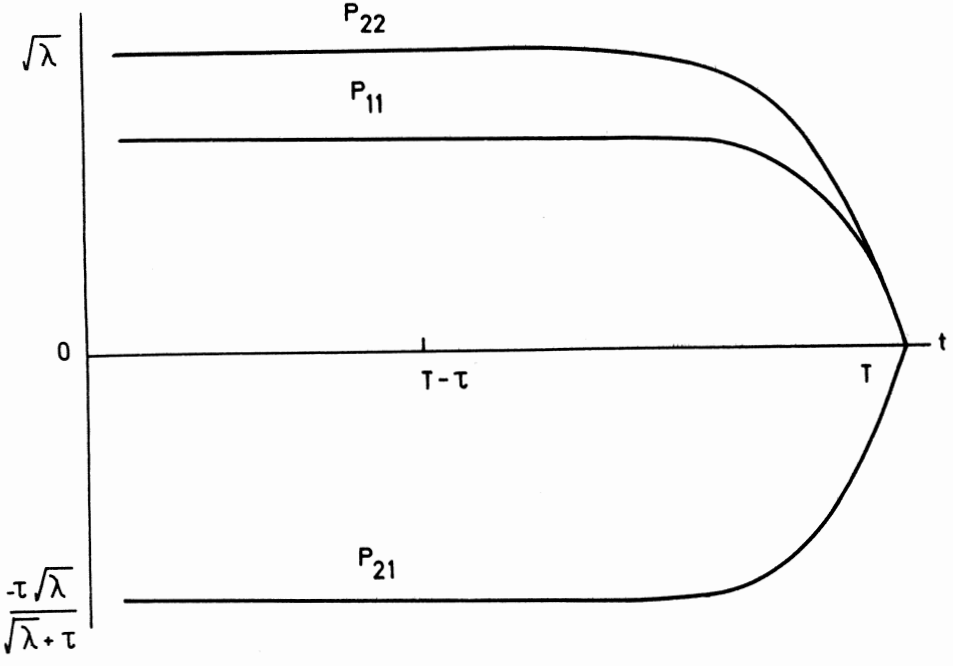


FIG 11.3-2 Solution of equations (11.3-3)

The optimal control is given by

$$u_{opt} = -\frac{1}{\lambda}\begin{bmatrix}0 & 1\end{bmatrix}\begin{bmatrix}p_{11} & p_{21}\\ p_{21} & p_{22}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

$$= -\frac{1}{\lambda}\left(p_{21}\,x_1 + p_{22}\,x_2\right)$$

If only the steady state situation is considered then the appropriate constant values for p_{21} and p_{22} can be obtained directly from equations (11.3-3b) and (11.3-3c) by putting $\dot{p}_{21} = \dot{p}_{22} = 0$ without the necessity to integrate all three equations (11.3-3).

Thus
$$0 = -1 - \frac{p_{21}}{\tau} - \frac{p_{21}^2}{\lambda}$$

$$0 = 1 - \frac{p_{22}^2}{\lambda}$$

from which
$$p_{22} = \sqrt{\lambda}$$

and
$$p_{21} = \frac{-\tau\sqrt{\lambda}}{(\sqrt{\lambda} + \tau)}$$

Hence, for a long time to go before interception

$$u_{opt} = \frac{\tau}{\sqrt{\lambda}\,(\sqrt{\lambda} + \tau)}\, x_1 - \frac{1}{\sqrt{\lambda}}\, x_2$$

This control can be implemented as shown in Figure 11.3-3 provided $x_1(\theta_T)$ and $x_2(\theta_M)$ can be measured or otherwise determined.

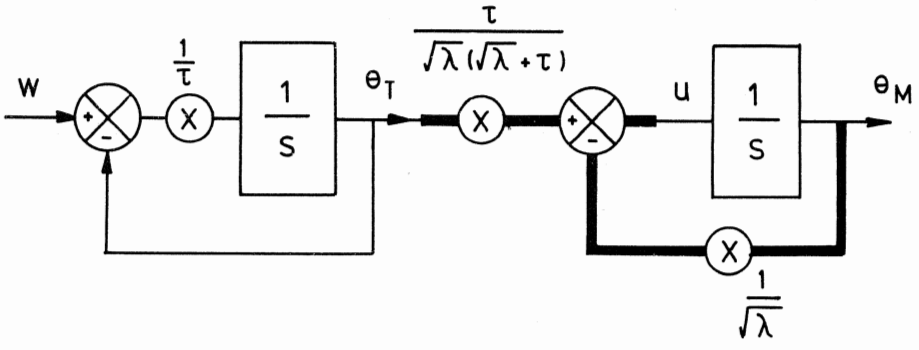


FIG 11.3-3 Optimal controller for two state stochastic system

By redrawing this optimal controller (Figure 11.3-4) it can be seen that the final design comprises a closed loop system together with a feedforward path.

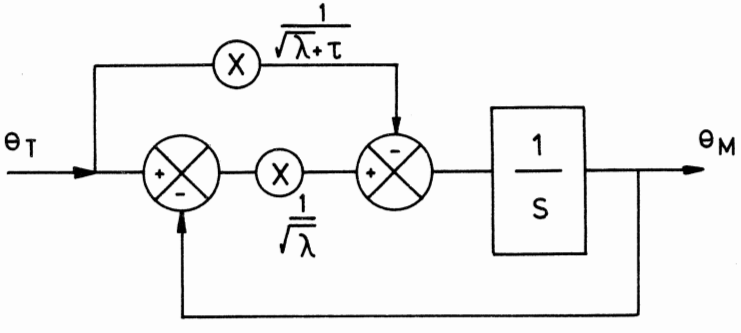


FIG 11.3-4 Alternative representation of optimal controller

By letting $\lambda \to 0$ (i.e. no constraint on the control "power") and noting from the target motion model that, on average, $\dot{\theta}_T = -\frac{1}{\tau}\theta_T$ then it is seen that the feedforward term is effectively a measure of $\dot{\theta}_T$. In other words, "classical" feedforward together with a negative feedback loop, of bandwidth $(\frac{1}{\sqrt{\lambda}})$ determined by the available control power, has emerged directly from the mathematics.

It is interesting to note that the transfer function relating θ_M to θ_T, viz

$$\frac{\theta_M}{\theta_T}(s) = \frac{\frac{\tau}{\tau + \sqrt{\lambda}}}{1 + \sqrt{\lambda}\, s}$$

is the same as that part of the Wiener filter solution, equation (10.4-1), which we associated with the need to constrain the control variable ($u = \dot{\theta}_M$). We have not reproduced the full Wiener filter solution, with the additional factor $\frac{1}{2 + \tau s}$, since we have assumed, in this chapter, that there is no measurement noise ($\theta_N = 0$).

Although the Wiener filter and optimal control approaches lead to the same overall solution, the greater insight gained into the underlying structure of the control system by the latter method is often of more use to the engineer than the values of the controller parameters. Also, for more realistic, and hence more complicated problems, the solution of the optimal control equations is usually far easier than the spectral factorisation required by the Wiener filter method. In fact the spectral factorisation problem has been exchanged for one of solving a system of non-linear algebraic equations. For high order systems it is usually easiest to solve these by integrating the Riccati equations backwards until the steady state solutions have been attained (see Figure 11.3-2).

It should be noted that the optimal control design procedure outlined in appendix A holds only for a linear system with a specific type of PI. Although many practical situations can be usefully reduced to this form the treatment of non-linear systems is far less straightforward. Nevertheless, by representing the system in state variable form it is still possible to use the computer to obtain, iteratively, a numerical solution to the control problem although an explicit feedback control system can seldom be produced in this case.

Before proceeding to the next chapter where the problem of estimating the states θ_T and θ_M in the presence of noise (θ_N) is considered we will apply the results of appendix A to the design of a homing system.

11.4 OPTIMAL CONTROL APPLIED TO A HOMING SYSTEM

The intention in this section is not only to give a further example of the application of optimal control theory but also to afford a complementary approach to the design of homing systems discussed previously in chapter 9.

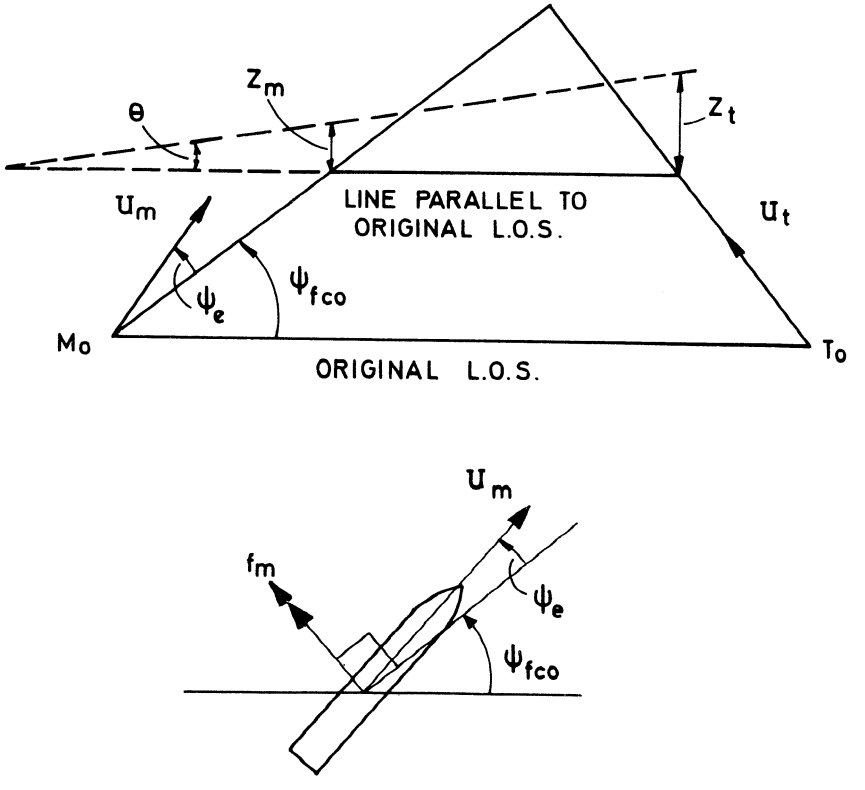


FIG 11.4-1 Interception geometry

Using the notation of chapter 9, repeated here in Figure 11.4-1, it would seem reasonable to choose as one state variable:

$$x_1 = z_t - z_m$$

since $z_t - z_m$ is a measure of the miss-distance which must obviously be included somewhere in the PI.

There will be one remaining state variable to choose since we are controlling missile acceleration and thus need two integrators, i.e. two states, in our model representing the dynamics of the missile system. Further state variables would be needed to specify the target motion if it were assumed to be other than constant velocity. The case of an accelerating target is considered by

Stallard (4) but, for simplicity, we will assume constant U_t.
The remaining state variable is conveniently taken to be:

$$\begin{aligned} x_2 &= \dot{x}_1 \\ &= \dot{z}_t - \dot{z}_m \\ &= -\dot{z}_m \quad \text{since } U_t \text{ is constant} \end{aligned}$$

$$\text{therefore } \dot{x}_2 = -\ddot{z}_m = -u$$

where u is the demanded acceleration normal to the original line of sight ($M_o\ T_o$) assuming a zero lag autopilot, that ψ_e is small and that missile incidence can be neglected completely.
In vector-matrix form the system is thus represented by:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u \qquad (11.4\text{-}1)$$

Following the usual procedure we will define a PI and then find the control u that will minimise it. This PI must be of quadratic form if the results of appendix A are to be used. However, there is no reason why a weighting on the final values of selected states should not be included in the PI provided it satisfies the form given by equation (A2). For the CLOS systems considered so far we have ignored this extra term in the PI (by taking $P(T) = 0$ in equation (A2)) but in order to provide a comparison with results obtained in chapter 9 we will choose a PI

$$J = x_1^2(T) + \lambda \int_o^T u^2\, dt \qquad (11.4\text{-}2)$$

The first term represents the requirement to minimise the miss at interception and the integral term represents the requirement to minimise energy expenditure, in the form of induced drag effects caused by manoeuvre, throughout the flight.
Comparing the chosen PI (equation (11.4-2)) with the standard form (equation (A2)) we have

$$P(T) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ Q = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},\ R = \lambda$$

Also, from equation (11.4-1), we have

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

By substituting for A, B, Q and R in the Riccati equation (A10) and expanding, the problem becomes one of solving the three non-linear differential equations.

$$-\dot{p}_{11} = -\frac{p_{21}^2}{\lambda} \tag{11.4-3a}$$

$$-\dot{p}_{21} = p_{11} - \frac{p_{21}p_{22}}{\lambda} \tag{11.4-3b}$$

$$-\dot{p}_{22} = 2p_{21} - \frac{p_{22}}{\lambda} \tag{11.4-3c}$$

subject to the boundary conditions $p_{11}(T) = 1$, $p_{21}(T) = 0$, $p_{22}(T) = 0$. These equations have been solved numerically by integrating backwards from time T and the results are shown in Figure 11.4-2 for a particular λ.

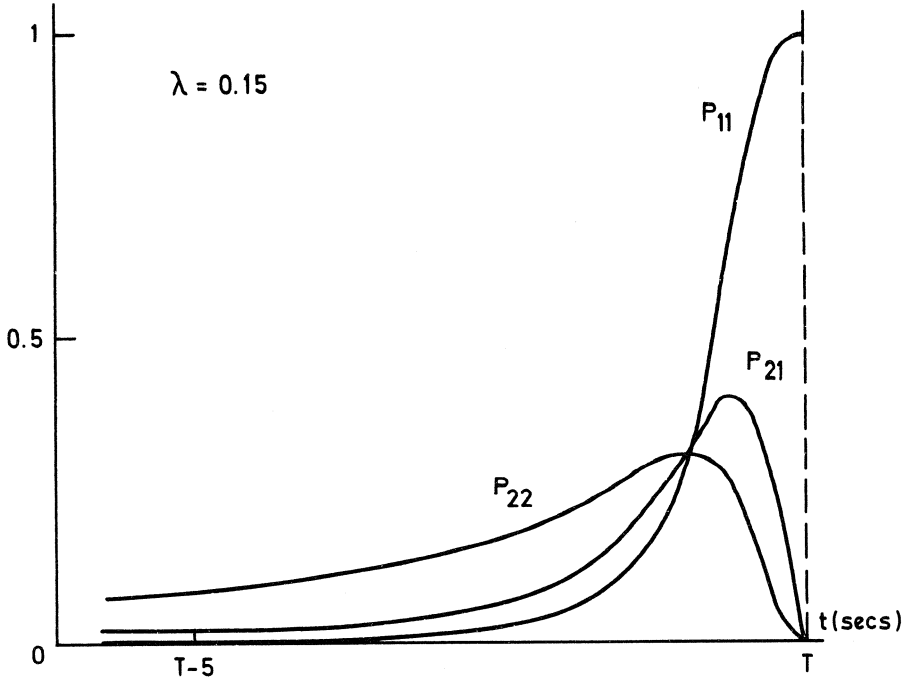


FIG 11.4-2 Solution of equations (11.4-3)

It is interesting to compare these results with, for example, those of Figure 11.3-2 where no weighting on the final states was included in the PI. Not unexpectedly, in this present case no significant control is applied to the missile until the interception time is approached when p_{11}, p_{21} and p_{22} undergo rapid changes. Of course, the optimum variation in these parameters can only be realised if the interception time T is known. If the homing system uses a pulse radar it would be feasible to extract an estimate of T but

whether the benefits obtained from optimal gain variations compensate for the extra complexity required to vary those gains and in addition to estimate T would be an important practical consideration.

Although not generally available in problems of this type, analytical solutions do happen to exist for equations (11.4-3), viz

$$p_{11} = \frac{3\lambda}{3\lambda + (T - t)^3} = f(t)$$

$$p_{21} = (T - t)\, f(t)$$

$$p_{22} = (T - t)^2\, f(t)$$

and these can readily be verified by substitution. Now from equation (9.3-2)

$$\theta \simeq \frac{(z_t - z_m)}{U_r\, \tau} = \frac{x_1}{U_r\, (T - t)}$$

where U_r is the closing velocity and τ is the time to go to interception. Therefore, to a close approximation if θ is small,

$$x_1 = U_r\, (T - t)\, \theta$$

$$\text{and } x_2 = U_r\, ((T - t)\, \dot{\theta} - \theta)$$

or, in vector-matrix form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = U_r \begin{bmatrix} (T - t) & 0 \\ -1 & T - t \end{bmatrix} \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} \qquad (11.4\text{-}4)$$

The optimal control is therefore given by

$$u_{opt} = -R^{-1} B^T P \underline{x}$$

$$= (T - t)^3\, f(t)\, U_r\, \dot{\theta}/\lambda \qquad (11.4\text{-}5)$$

It was shown in chapter 9 that the acceleration normal to the original line of sight in a classical proportional navigation system is given by

$$u = f_m \cos \psi_{fco}$$

$$= \alpha\, U_r\, \dot{\theta} \qquad (11.4\text{-}6)$$

where α is the kinematic gain.

Comparing equations (11.4-5) and (11.4-6) it is seen that optimal control is also proportional navigation but with a time varying kinematic gain

$$\alpha = (T - t)^3\, f(t)/\lambda$$

$$= \frac{3(T - t)^3}{3\lambda + (T - t)^3}$$

It is tempting to conclude that if $\lambda = 0$, implying a PI

$$J = x_1^2 \ (T) \qquad (11.4\text{-}7)$$

then the optimal control is given by $\alpha = 3$. However, the Matrix-Riccati equation (A10) cannot be applied if $\lambda = 0$ since $R^{-1} = \lambda^{-1}$ is undefined. Hammond (5) has studied the problem associated with the minimisation of J (equation 11.4-7) and concludes that there is an infinite number of controllers that would result in $J = 0$.

This rather unhelpful result, from a design point of view, simply emphasises how important it is to choose a realistic PI. In any practical situation account must be taken of the energy available to accomplish the task and the most convenient way of doing this is to include a measure of u in the PI. Finally, let us consider the PI

$$J = \int_0^T u^2 \, dt$$

The optimal control law is derived exactly as previously except that now $\lambda = 1$ and $P(T) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

It can be shown, by substitution, that

$$p_{11} = \frac{3}{(T-t)^3}, \ p_{21} = \frac{3}{(T-t)^2}, \ p_{22} = \frac{3}{(T-t)}$$

are solutions of the Matrix-Riccati equation and, on forming the optimal control, we get

$$u_{opt} = \frac{3}{(T-t)^2} x_1 + \frac{3}{(T-t)} x_2$$

or, using equation (11.4-4)

$$u_{opt} = 3 \, U_r \, \dot{\theta}$$

which agrees with the result obtained in chapter 9 (equation 9.5-10).

REFERENCES

1. PITMAN D.L. Optimisation and Kalman filter. Paper 3c. AGARD lecture series No 52 May 1972.

2. GARDNER A.H. A simulation study of optimal and sub-optimal control of second and fourth order dynamical systems. RAE Tech Memo WE1399 1974.

3. JACOBS O.L.R. Introduction to control theory. O.U.P. 1974.

4. STALLARD D.V. Classical and modern guidance of homing interceptor missiles. Raytheon Company Report P247 April 1968.

5. HAMMOND J.K. Proportional Navigation - a singular problem in optimal control. A.A.S.U. Report No 309 September 1971.
6. ANDERSON B.D.O. and J. B. MOORE Linear Optimal Control. Prentice-Hall 1971.

CHAPTER 12

KALMAN FILTERS

12.1 PROBLEM REVIEW

In chapters 10 and 11 we tackled the design of an optimal stochastic control system, i.e. the minimisation of a PI when the future values of the state variables are not well defined but are random functions of time. In order that some progress could be made we restricted ourselves to stationary random signals and assumed that their power spectra were available.

Also, in chapter 10, this problem was tackled under the handicap that even the present value of the state θ_T (the angular position of the target being tracked) could not be measured exactly but was contaminated with noise θ_N.

The technique used, Wiener filter theory, led to a solution (equation (10.4-1)) which represented the overall physically realisable transfer function F_{pr} relating the measurement $(\theta_T + \theta_N)$ to the missile position θ_M (Figure 12.1-1).

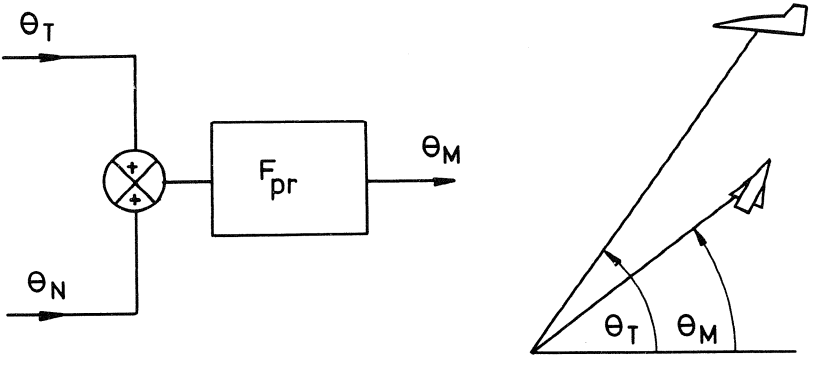


FIG 12.1-1 Optimum physically realisable filter F_{pr}

Only from experience with previous examples and anticipation of the results of chapter 11 was it possible to identify the factors of F_{pr} with the separate functions of "control" and "filtering". Furthermore, a closed loop structure for F_{pr} had still to be extracted before the design could be implemented. This, together with the need for spectral factorisation in the derivation of F_{pr}, has made the method unpopular with engineers for whom a knowledge of system structure is of paramount importance.

The techniques of modern control theory were used in section 11.3 where the same stochastic control problem was tackled although, as a temporary expedient, it was assumed that the measurement of θ_T was not contaminated with noise. It was seen that a feedback control law resulted directly and thus a system structure emerged which could be used for subsequent design studies. Also, it was noted that for a long time to go to interception, when the optimal control gains remain constant, the overall transfer function so obtained corresponded to the "control" function part of the Wiener filter.

This chapter will consider the means for providing the optimal controller in section 11.3 with information on the system states, x_1 (θ_T) and x_2 (θ_M), when the measurement of one of them (θ_T) is contaminated with noise (θ_N). This will complete the problem originally posed and solved using Wiener filter theory, in *Example 10.4-1*.

More generally, we require a device that will estimate *all* the system states, given inaccurate measurements of a few (perhaps only one) of them. This is the role that the Kalman filter fulfills.

12.2 INTRODUCTION TO THE KALMAN FILTER

Consider the two state stochastic system of section 11.3, repeated here in Figure 12.2-1.

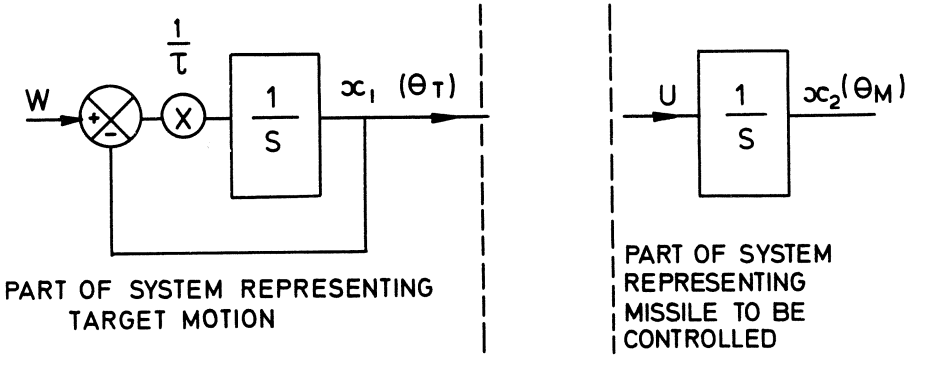


FIG 12.2-1 Two state stochastic system

Figure 12.2-1 represents a mathematical model of the actual system. If an equivalent physical model could be constructed on some form of simulator then, in principle at least, it would be possible to use the values of θ_T and θ_M given by the simulator model without any further reference to the original system. Provided the simulator model is started with the correct initial conditions it will continue to mimic the behaviour of the actual system, thus

eliminating the need for measurements on that system.

Unfortunately such an ideal situation cannot obtain. Even if the actual system model structure, parameter values and initial conditions are known exactly the input w, being a random function of time, cannot be reproduced in the simulator model. Effectively w represents the uncertainty in our knowledge of the actual system.

Of course, in practice some measurements will be made on the system. These measurements will not be perfect since there are always errors associated with any measurement process and we recognised this fact in *Example 10.4-1* where we assumed a white measurement noise θ_N on θ_T. Any measurement of θ_M, subsequently required in implementing F_{pr} as a negative feedback system, was tacitly assumed to be perfect.

However, it is usually the case that a CLOS tracking system will give a measure not of θ_T and θ_M separately but only the difference between them ($\theta_T - \theta_M$). This situation is depicted in Figure 12.2-2 together with a more general notation where z represents the measurement and v the measurement noise.

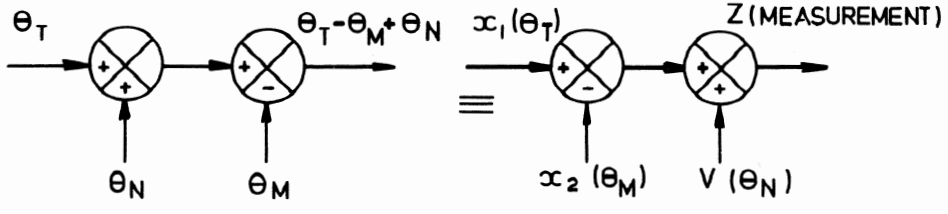


FIG 12.2-2 Measurement z contaminated by noise v

Our task is to obtain best estimates, in some sense, of the states x_1 and x_2 given the measurement z, a knowledge of the structure of the system and its parameter values, and the statistical characteristics of the noise sources w and v. The assumed system model, which is equivalent to the previously proposed simulator model, is identical to the actual system model except for the zero mean white noise source w. Since the best guess of w at any instant is zero it is omitted entirely and thus represents the uncertainty in the assumed system model. The problem is summarised in Figure 12.2-3.

The Kalman filter tackles this problem by treating the measurement process and system model outputs as two independent estimates of the state of the actual system. It combines these outputs together to form a best estimate (in the sense of having minimum variance).

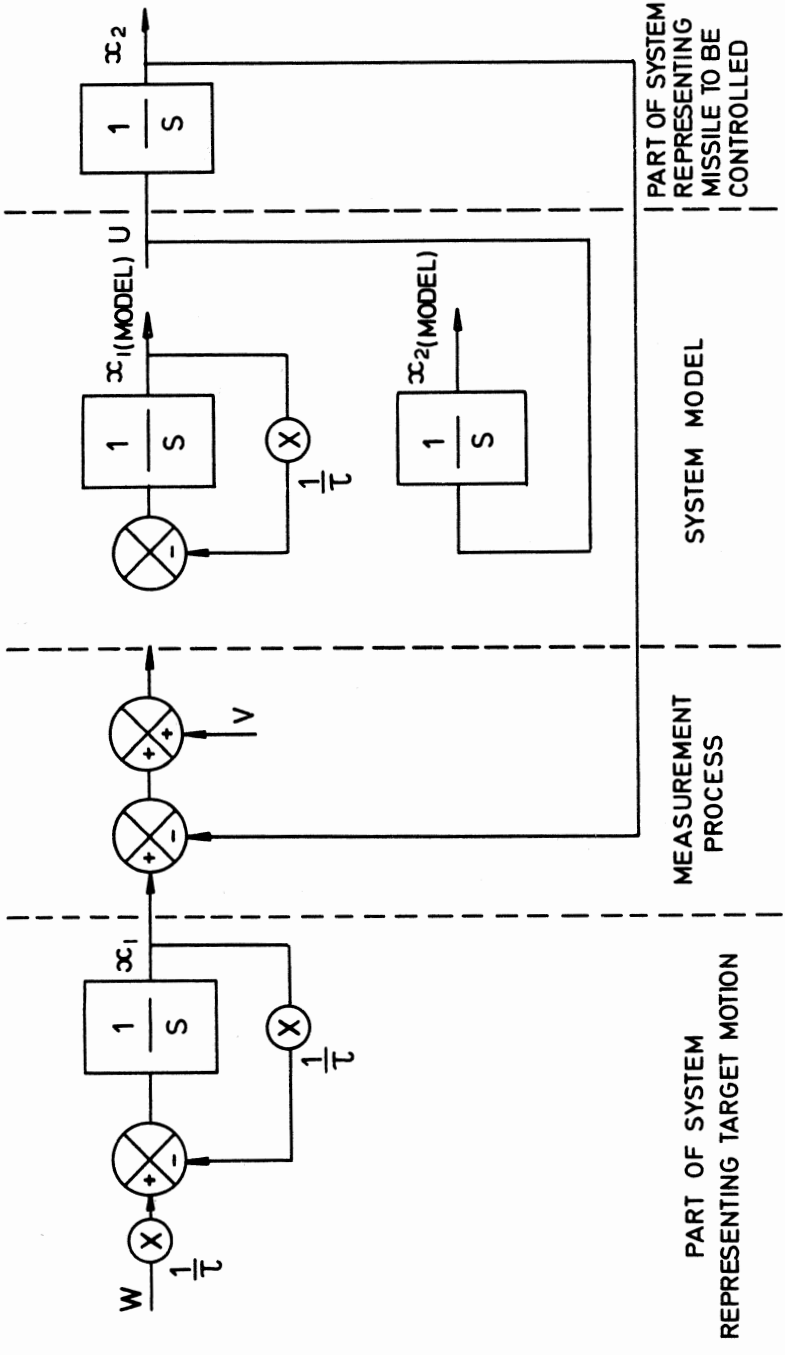


FIG 12.2-3 To obtain best estimates of x_1 and x_2 given a system model and a measurement z

12.3 THE DISCRETE KALMAN FILTER

Conceptually, the Kalman filter is most readily formulated in its discrete (or sampled) form. Consider the single dimensional continuous stochastic system (Figure 12.3-1) where A and B are time invariant and w is a zero mean white noise source.

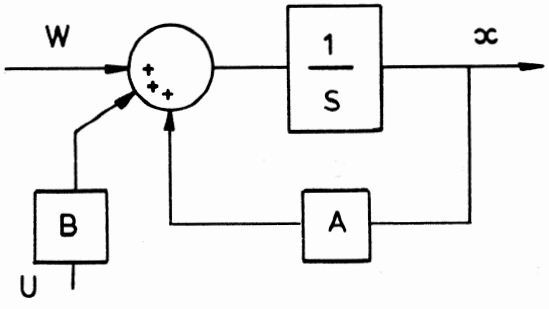


FIG 12.3-1 Single dimensional stochastic system

The equation describing this system is

$$\dot{x} = Ax + Bu + w \tag{12.3-1}$$

In discrete terms this can be written

$$\frac{X_{n+1} - X_n}{\Delta t} = AX_n + BU_n + W_n \tag{12.3-2}$$

where $X_n = x(t_n)$ etc and Δt is the sampling interval. Provided Δt is sufficiently small equation (12.3-2) is a close approximation to equation (12.3-1). Re-writing equation (12.3-2)

$$X_{n+1} = \Phi X_n + B\,\Delta t\,U_n + \Delta t\,W_n \tag{12.3-3}$$

where $\Phi = (1 + A\Delta t)$.

The assumed variance of W_n, a zero mean white noise sequence after sampling, is σ_w^2.

Suppose now that a measurement Z_{n+1} of X_{n+1} is made and that this measurement is contaminated by a white noise sequence V_{n+1} with zero mean and variance σ_v^2 as shown in Figure 12.3-2.

Besides the measurement Z_{n+1}, we will also assume that we have *a priori* knowledge of the system from which a model can be constructed. This model will be imperfect to the extent that the white noise sequence W_n in equation (12.3-3) is unknown. However, since the mean value of W_n is zero we cannot do better, with the available information, than to assume that W_n is zero in our model. The *a priori* model of the system can thus be constructed from equation (12.3-3) with $W_n = 0$.

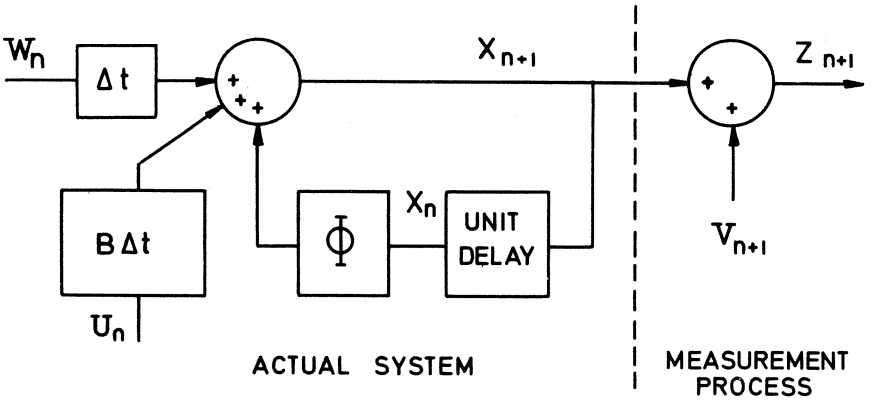


FIG 12.3-2 Single dimensional discrete stochastic system and measurement process

Putting the cart before the horse, let us assume that the required best estimate $\hat{X}_{n+1}$ of X_{n+1} is already available. If Y_{n+1} is our model prediction of X_{n+1} based on the best estimate $\hat{X}_n$ from the previous sampling interval, and if Z_{n+1} is the current measured value of X_{n+1} then we have the situation shown in Figure 12.3-3.

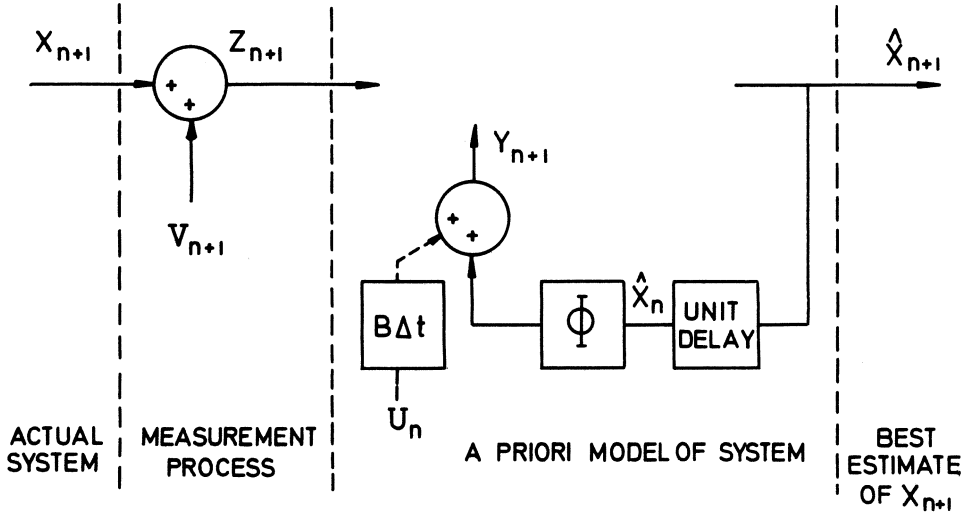


FIG 12.3-3 Discrete measurement process and a priori model

Taking the variance of Y_{n+1} to be $\sigma_{Yn+1}{}^2$, as yet undetermined, we now have available *two* independent estimates of X_{n+1}:

(i) the measurement Z_{n+1}, variance $\sigma_v{}^2$

(ii) the model prediction Y_{n+1}, variance $\sigma_{Y_{n+1}}{}^2$

Because U_n, the control input, is common to both the actual system and the model it can be ignored in the present discussion since it does not affect the accuracy of the model and therefore will not influence the choice of the best estimate of X_{n+1}. The model equation is therefore:

$$Y_{n+1} = \Phi \hat{X}_n \tag{12.3-4}$$

It can be shown (1) that the best (minimum variance) estimate $\hat{X}_{n+1}$ of X_{n+1} is given by a linear combination of Z_{n+1} and Y_{n+1} with weighting factor K_{n+1} such that

$$\hat{X}_{n+1} = (1 - K_{n+1})\, Y_{n+1} + K_{n+1}\, Z_{n+1} \tag{12.3-5}$$

$$\text{where } K_{n+1} = \frac{\sigma_{Y_{n+1}}^2}{\sigma_{Y_{n+1}}^2 + \sigma_v^2} \tag{12.3-6}$$

The physical interpretation of this result is straightforward. It implies that the more uncertain we are about Z_{n+1} (σ_v^2 large compared to $\sigma_{Y_{n+1}}^2$) then the more notice we take of Y_{n+1}, and vice versa.
The best estimator (or filter) is thus as shown in Figure 12.3-4.

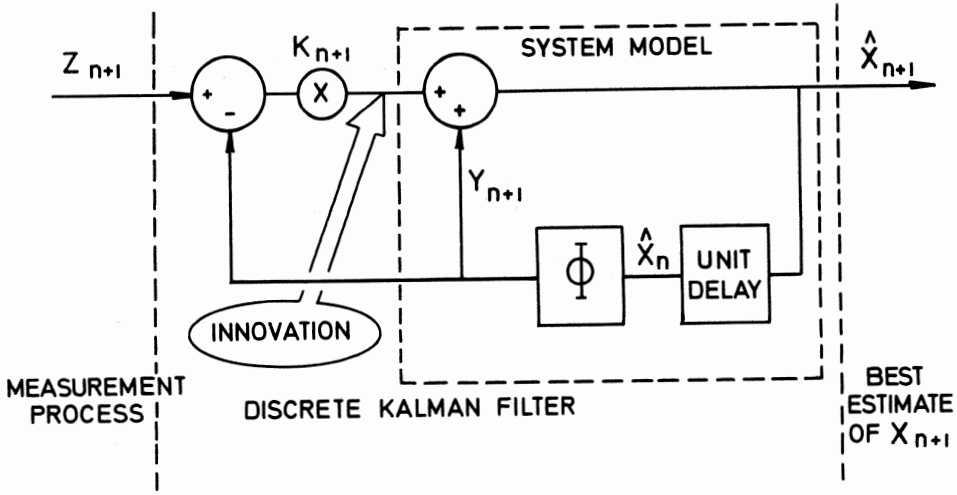


FIG 12.3-4 Discrete Kalman filter

It should be noted that the filter is essentially a model of the system under observation updated by a signal, often called the innovation or residual, proportional to the difference between the model and measured estimates of X_{n+1}. The only unknown is the value of $\sigma_{Y_{n+1}}^2$ needed to calculate the optimum weighting factor K_{n+1} (equation (12.3-6)). This can be found by making use of a standard

result from statistics which says that if two independent normal random variables are combined as shown in Figure 12.3-5 then the variance of the combination is given by

$$\sigma^2 = K_1^2 \sigma_1^2 + K_2^2 \sigma_2^2 \qquad (12.3\text{-}7)$$

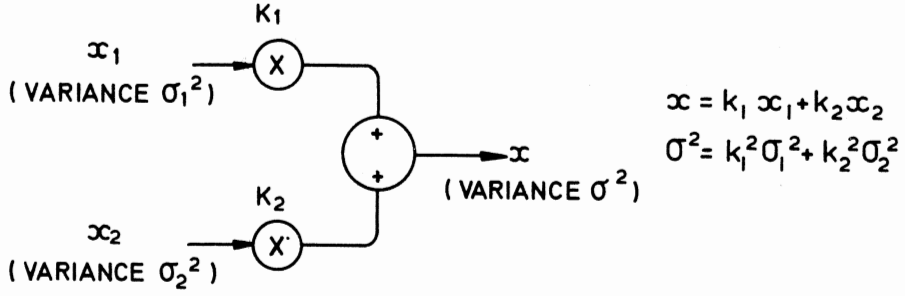


FIG 12.3-5 Linear combination of normal random variables

Hence we can say that

$$\sigma_{Y_{n+1}}^2 = \Phi^2 \hat{\sigma}_n^2 + \Delta t^2 \sigma_w^2 \qquad (12.3\text{-}8)$$

In other words, the uncertainty in the model output, compared to the true state X_n, is due to two factors (see Figure 12.3-6).

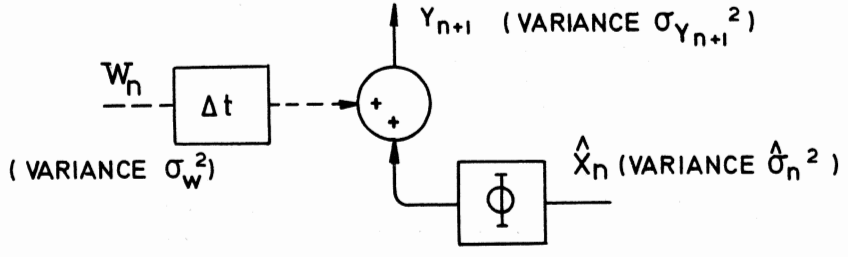


FIG 12.3-6 Causes of uncertainty in Y_{n+1}

Firstly, the model is driven by the best estimate $\hat{X}_n$ and not by the true state X_n.

Secondly, the *absence* of the input W_n which is present in the real system but not in the model and which therefore represents the uncertainty in the model itself.

The variance or uncertainty in $\hat{X}_n$ is represented by $\hat{\sigma}_n^2$ and this too must be calculated. Fortunately, it can be deduced immediately from equations (12.3-5) and (12.3-7) as

$$\hat{\sigma}_n^2 = (1 - K_n)^2 \sigma_{Y_n}^2 + K_n^2 \sigma_v^2$$

and making use of equation (12.3-6) to eliminate σ_v^2 gives

$$\hat{\sigma}_n^{\,2} = (1 - K_n)\,\sigma_{Y_n}^{\,2} \qquad (12.3\text{-}9)$$

Finally, collecting together equations (12.3-4) to (12.3-9), omitting equation (12.3-7), we have the complete set of discrete Kalman filter equations for a single dimensional system:

$$\begin{aligned}
Y_{n+1} &= \Phi\,\hat{X}_n && \text{model estimate of } X_{n+1}\\
\sigma_{Y_{n+1}}^{\,2} &= \Phi^2\,\hat{\sigma}_n^{\,2} + \Delta t^2\,\sigma_w^{\,2} && \text{variance of } Y_{n+1}\\
K_{n+1} &= \frac{\sigma_{Y_{n+1}}^{\,2}}{\sigma_{Y_{n+1}}^{\,2} + \sigma_v^{\,2}} && \text{filter gain}\\
\hat{X}_{n+1} &= Y_{n+1} + K_{n+1}\,(Z_{n+1} - Y_{n+1}) && \text{best estimate of } X_{n+1}\\
\hat{\sigma}_{n+1}^{\,2} &= (1 - K_{n+1})\,\sigma_{Y_{n+1}}^{\,2} && \text{variance of } \hat{X}_{n+1}
\end{aligned} \qquad (12.3\text{-}10)$$

It should be noted that some derivations of these equations take W_n to be a noise sequence operating directly on Y_{n+1} and in that case the coefficient of $\sigma_w^{\,2}$ in equation (12.3-8) is unity and not Δt^2.

In the multi-dimensional case Y_{n+1}, $\hat{X}_n$ etc become vectors, Φ is a matrix and $\sigma_{Y_{n+1}}^{\,2}$, $\hat{\sigma}_{n+1}^{\,2}$ etc are represented by covariance matrices but the form of the equations remains unaltered. Provision must also be made for the fact that, in general, the number of measurements will be less than the number of states and a matrix multiplication must be performed on the model outputs in order to make dimensions compatible before forming the innovation. The reader interested in a more detailed treatment of the multi-dimensional Kalman filter is strongly recommended to see Barham and Humphries (1).

It should be noted that in order to produce best estimates in the $(n+1)^{\text{th}}$ interval, the five equations (12.3-10) only make use of information from the n^{th} interval. They constitute a recursive algorithm with modest storage requirements since all past measurement data is discarded. They are, therefore, ideally suited to implementation on a digital computer and indeed this is how most Kalman filtering is done.

However, in order to demonstrate the remarkable similarity between optimal control (as developed for continuous systems in chapter 11) and optimal estimation (the Kalman filter developed above) it will be necessary to consider the continuous form of the Kalman filter.

12.4 THE CONTINUOUS KALMAN FILTER

The continuous Kalman filter equations can be formally derived from equations (12.3-10) by taking the limit as the sampling interval Δt tends to zero. Initially, however, we will adopt an approach based largely on a physical reinterpretation of the results obtained for the discrete Kalman filter. From our experience with the discrete Kalman filter we can argue that the continuous Kalman filter must contain a model of the system whose states we are trying to estimate. Also, this model will be continuously driven by an innovation signal, proportional to the difference between the model and measured values. This situation is depicted in Figure 12.4-1.

We can also argue that the filter gain (K) must be such as to provide an innovation dependent on the relative uncertainties in the measured and model values. However, it should be noted that an essential difference between the continuous and discrete Kalman filter is that in the continuous case the model output is also the best estimate.

Hence, using equations (12.3-6) and (12.3-9) in order to express K_{n+1} in terms of the variance of the best estimate ($\hat{\sigma}_{n+1}^2$) rather than the model variance ($\sigma_{Y_{n+1}}^2$) we get

$$K_{n+1} = \frac{\hat{\sigma}_{n+1}^2}{\sigma_v^2} \tag{12.4-1}$$

as an alternative form for the discrete Kalman filter gain.

In order to derive K for the continuous Kalman filter we must consider how equation (12.4-1) behaves as $\Delta t \to 0$. From equations (12.3-10) we have

$$\begin{aligned}\hat{X}_{n+1} &= Y_{n+1} + K_{n+1}\,(Z_{n+1} - Y_{n+1}) \\ &= \Phi\,\hat{X}_n + K_{n+1}\,(Z_{n+1} - \Phi\,\hat{X}_n)\end{aligned}$$

Replacing Φ by $(1 + A\Delta t)$ (see equation (12.3-3)) and rearranging gives

$$\frac{\hat{X}_{n+1} - \hat{X}_n}{\Delta t} = A\,\hat{X}_n + \frac{K_{n+1}}{\Delta t}\,(Z_{n+1} - \hat{X}_n - A\,\Delta t\,\hat{X}_n)$$

Hence, taking the limit as $\Delta t \to 0$, we get

$$\dot{\hat{x}} = A\,\hat{x} + K\,(z - \hat{x}) \tag{12.4-2}$$

where K is defined by the limit

$$\frac{K_{n+1}}{\Delta t} \to K \text{ as } \Delta t \to 0 \tag{12.4-3}$$

Equation (12.4-2) confirms the form of the continuous Kalman filter deduced previously (Figure 12.4-1).

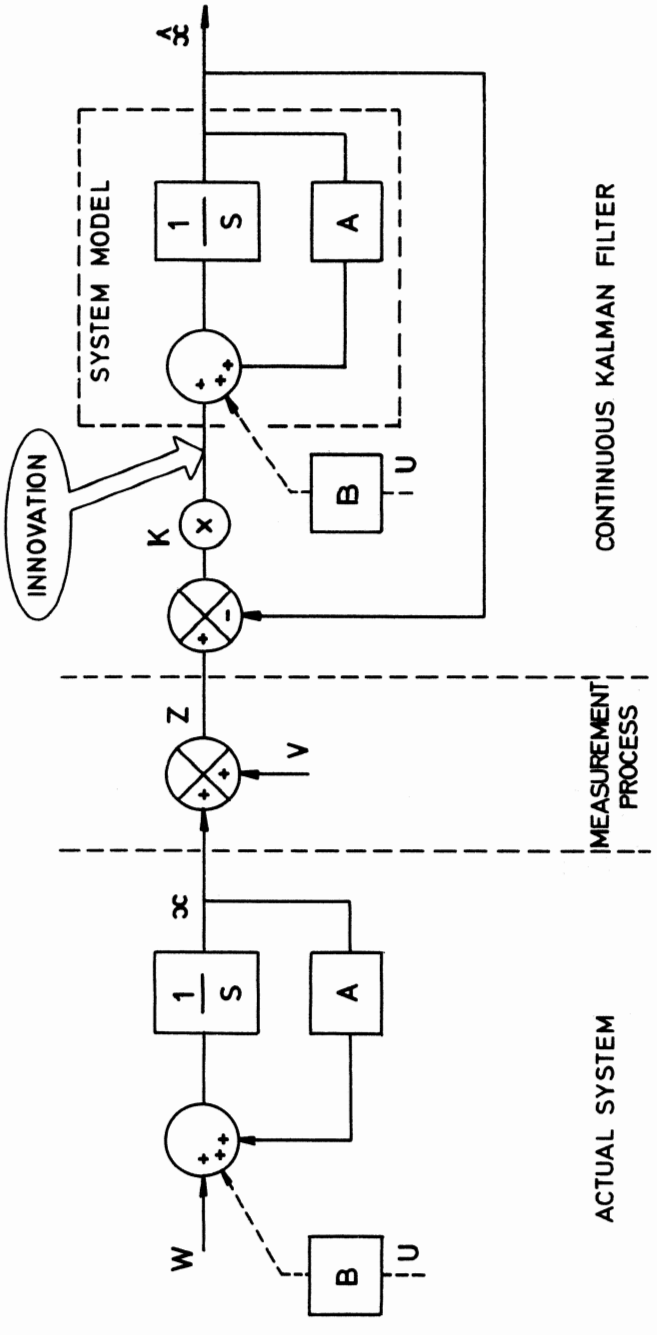


FIG 12.4-1 Single dimensional continuous Kalman filter

However, equations (12.4-1) and (12.4-3) imply that K is the limit of $\frac{\hat{\sigma}_{n+1}^2}{\sigma_v^2}\Delta t$ as $\Delta t \to 0$.

This does not mean that K is indeterminate but rather that we must look more closely at the way the other terms in the expression behave as the sampling time $\Delta t \to 0$. In particular, we must consider in more detail how the white noise sequences w_n and v_n behave in this situation.

A sequence is just a succession of numbers, "white" indicating that each number is unrelated to or uncorrelated with the previous number of the sequence. The variance is simply a measure of the spread of these successive values about the average, in this case zero. As $\Delta t \to 0$ the white noise sequence becomes a continuous white noise process and the variance, which can now be interpreted as the "power" of the random process, tends to infinity. This is always a problem when dealing with fictional infinite bandwidth noise and is overcome by describing the statistical characteristics of the white noise in terms of its spectral density, i.e. power per unit bandwidth.

The form of the relationship between the variance of the white noise sequence and the spectral density of the white noise process can be demonstrated if the white noise sequence is considered to be the result of filtering the continuous white noise process over the sampling interval (Figure 12.4-2).

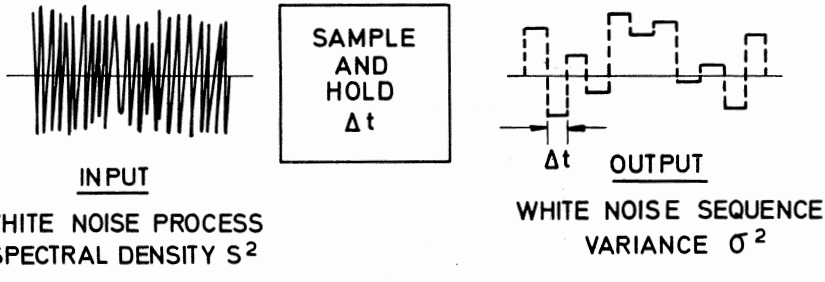


FIG 12.4-2 Sampled white noise

It is seen that the variance (σ^2) of the white noise sequence, considered as the power content of the white noise process after filtering, is, in some way, inversely proportional to the sampling interval. Clearly, as we increase the sample time Δt (crudely equated to the time constant of a low pass filter), more and more of the high frequency power of the input white noise process will be eliminated from the output white noise sequence.

In fact, it can be shown that

$$\sigma^2 \propto \frac{1}{\Delta t} \text{ (approximately)}$$

$$\text{and } \sigma^2 \to \frac{s^2}{\Delta t} \text{ as } \Delta t \to 0$$

where s^2 is the spectral density of the continuous white noise process. Hence, if we let the spectral densities of the system noise (w) and the measurement noise (v) in the continuous Kalman filtering situation (Figure 12.4-1) be q^2 and r^2 respectively then:

$$\sigma_w^2 \to \frac{q^2}{\Delta t} \tag{12.4-4}$$

$$\text{and } \sigma_v^2 \to \frac{r^2}{\Delta t} \text{ as } \Delta t \to 0 \tag{12.4-5}$$

It should be noted that there is no such problem associated with the variance ($\hat{\sigma}_{n+1}^2$) of the best estimate since $\hat{X}_{n+1}$ does not form a white noise sequence. This is due to the inherent nature of the filter which, because of its imbedded model, ensures that there must be some correlation between successive best estimates.

Hence, from equations (12.4-1), (12.4-3) and (12.4-5) we have

$$K = \frac{\hat{\sigma}^2}{r^2} \tag{12.4-6}$$

as the limit as $\Delta t \to 0$.

It only remains to determine $\hat{\sigma}^2$ in order to establish the value of K and thus complete the derivation of the continuous Kalman filter.

From equations (12.3-8), (12.3-9) and (12.4-1) we have

$$\hat{\sigma}_{n+1}^2 = (1 - \frac{\hat{\sigma}_{n+1}^2}{\sigma_v^2})(\Phi^2 \hat{\sigma}_n^2 + \Delta t^2 \sigma_w^2)$$

Substituting for σ_w^2 and σ_v^2 from equations (12.4-4) and (12.4-5) gives

$$\hat{\sigma}_{n+1}^2 = (1 - \frac{\hat{\sigma}_{n+1}^2}{r^2} \Delta t)(\Phi^2 \hat{\sigma}_n^2 + q^2 \Delta t)$$

Finally, replacing Φ by $(1 + A\Delta t)$ and expanding, ignoring all terms of order Δt^2 or higher, we have

$$\hat{\sigma}_{n+1}^2 \simeq \hat{\sigma}_n^2 + 2A \Delta t \hat{\sigma}_n^2 + q^2 \Delta t - \frac{\hat{\sigma}_{n+1}^2 \hat{\sigma}_n^2}{r^2} \Delta t$$

$$\text{therefore } \frac{\hat{\sigma}_{n+1}^2 - \hat{\sigma}_n^2}{\Delta t} \simeq 2A \hat{\sigma}_n^2 + q^2 - \frac{\hat{\sigma}_{n+1}^2 \hat{\sigma}_n^2}{r^2}$$

In the limit as $\Delta t \to 0$, $\hat{\sigma}_{n+1}^2 \to \hat{\sigma}_n^2 \to \hat{\sigma}^2$ and

$$\dot{\hat{\sigma}}^2 = 2A\,\hat{\sigma}^2 + q^2 - \frac{(\hat{\sigma}^2)^2}{r^2} \qquad (12.4\text{-}7)$$

Thus the derivation of the Kalman filter gain requires the solution of a non-linear differential equation in $\hat{\sigma}^2$. This equation may be interpreted as follows.

The rate of change in the uncertainty ($\dot{\hat{\sigma}}^2$) in the best estimate is due to:

(a) A term $2A\,\hat{\sigma}^2$ which represents the isolated behaviour of the embedded system model within the filter.

(b) A term q^2 which represents the rate of *increase* of uncertainty due to system noise. This term is effectively an inverse measure of the "goodness" of the embedded model.

(c) A term $\frac{(\hat{\sigma}^2)^2}{r^2}$ which represents the rate of *decrease* of uncertainty due to incoming measurements.

It is interesting to compare the above results for the continuous Kalman filter with those derived in section 11.2 for continuous optimal control. In both cases a very similar non-linear differential equation must be solved in order to provide the optimum control or filter gain. This is just one example of the duality that exists between optimal control and estimation and it becomes even more apparent in the case of multivariable systems. Indeed, having established the form of the single dimensional continuous Kalman filter, we will now consider the multi-dimensional filter equations.

12.5 THE MULTI-DIMENSIONAL KALMAN FILTER

We will state, rather than derive, the necessary equations and will draw comparisons, whenever possible, with the results obtained previously. The reader requiring a detailed derivation is referred to Bryson and Ho (2).

The only essential changes that occur in going from single to multi-dimensional filters is that scalars become vectors and gains, variances and spectral densities are represented in matrix form. Provision must also be made for the case when the number of measurements is less than the number of states. This is achieved by the matrix H in Figure 12.5-1. The under bar notation (e.g. $\underline{x}$) indicates vector quantities and the use of tram line interconnections emphasises that this is a multidimensional estimator.

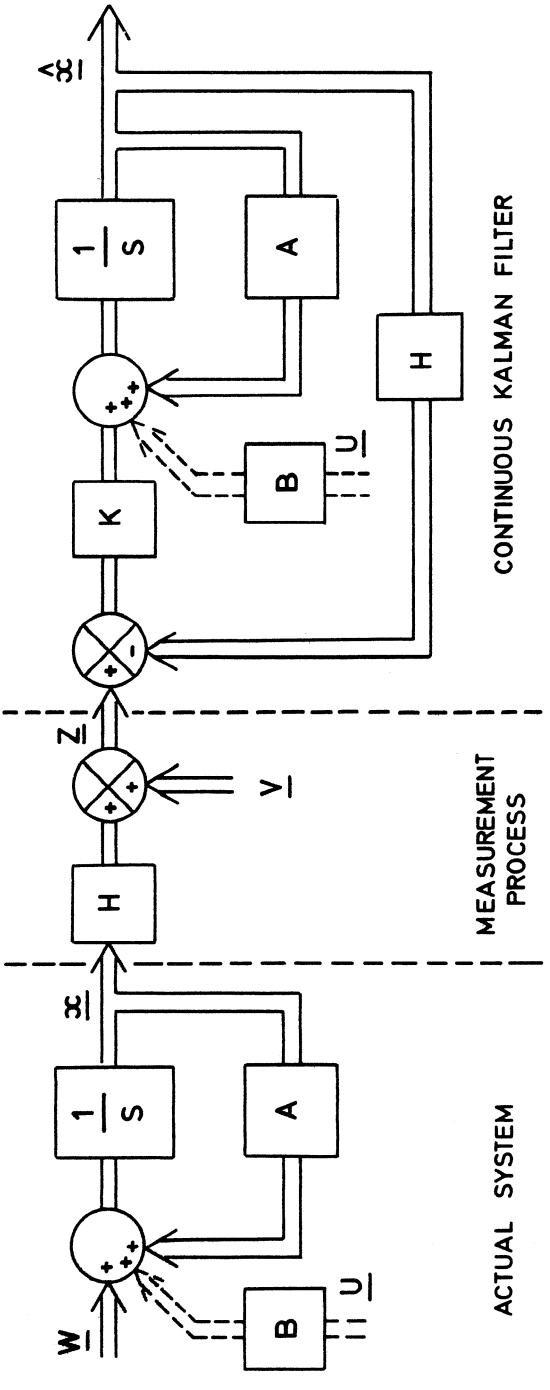


FIG 12.5-1 Multidimensional continuous Kalman filter

Let P, Q and R be the matrices corresponding to $\hat{\sigma}^2$, q^2 and r^2 respectively in the single dimensional filter. For the Q and R matrices, the elements on the main diagonal are the spectral densities of the individual system and measurement noise sources. Elements off the main diagonal indicate correlation between noise sources in terms of a cross spectral density but in many problems the noise sources are independent and hence these elements will be zero.

However, this is not the case for the P matrix. Here the off diagonal terms are covariances which indicate how one best estimate is related to another and these, in general, will not be zero. For this reason it is usual to refer to P as the covariance matrix even though the terms on the main diagonal still correspond to the variances of the individual best estimates.

The structural equations of the filter can, of course, be written down immediately from Figure 12.5-1. It is, however, the equation needed to calculate K, the filter gain matrix which is of most interest. In fact, it can be shown (2) that

$$K = PH^T R^{-1} \tag{12.5-1}$$

where P is given by the solution of the equation

$$\dot{P} = AP + PA^T + Q - PH^T R^{-1} HP \tag{12.5-2}$$

These equations are less forbidding if they are compared with the corresponding equations derived for the single dimensional Kalman filter in the previous section. Indeed, if the measurement matrix H is the identity matrix, simply implying that the number of measurements is equal to the number of states, then the comparison is complete. This is demonstrated below where the corresponding equations are listed:

Single dimensional

$$K = \frac{\hat{\sigma}^2}{r^2} \text{ where } \dot{\hat{\sigma}}^2 = 2A\hat{\sigma}^2 + q^2 - \frac{(\hat{\sigma}^2)^2}{r^2}$$

Multidimensional

$$K = PR^{-1} \text{ where } \dot{P} = AP + PA^T + Q - PR^{-1}P$$

Furthermore, comparison of equations (12.5-1) and (12.5-2) with equations (A9) and (A10) reveals that the optimal estimation process is very closely related to the optimal controller developed in chapter 11 and that a differential equation of the Riccati form is common to both. This should not be too surprising since in both cases the object is to minimise a quadratic performance index. For optimal control this is explicitly stated and for optimal estimation it is implicit in the requirement to find a best estimate, i.e. one with

minimum variance. Indeed, the optimal estimation equations can be deduced directly from appendix A by reformulating the estimation problem as a control problem, e.g. Anderson and Moore (3). The main difference between the two problems is that in the control case the boundary conditions for the Riccati equations are defined at interception and the equations are solved by integrating *backwards* in time. In the estimation case the boundary conditions are defined at the start of the engagement and represent initial guesses at the entries in the covariance matrix P. Provided the system is time invariant and the noise sources stationary then the elements of P should tend to steady values as the estimation process proceeds and these values can be found by integrating the Riccati equations *forward* in time.

For reference purposes the optimal estimator (Kalman filter) equations for a multidimensional continuous system are stated in appendix B. We will now make use of these equations to complete the problem posed in section 12.2 and summarised in Figure 12.2-3, viz given a noisy measurement of the error $(\theta_T - \theta_M)$, to obtain best (minimum variance) estimates of θ_T and θ_M. For convenience, we have re-stated the problem in Figure 12.5-2. We will assume that the system noise w has spectral density q^2 and the measurement noise v has spectral density r^2.

Thus, we have a system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\tau} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u + \begin{bmatrix} \frac{w}{\tau} \\ 0 \end{bmatrix} \qquad (12.5\text{-}3)$$

(cf equation (B1), appendix B)

and a measurement

$$z = (x_1 - x_2) + v$$

$$= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v \qquad (12.5\text{-}4)$$

(cf equation (B2), appendix B)

There is noise $\frac{w}{\tau}$ (spectral density $\frac{q^2}{\tau^2}$) on $\dot{x}_1$ but none on $\dot{x}_2$. Also, there is only one measurement and this is contaminated directly with the white noise v.

Hence, $$Q = \begin{bmatrix} \frac{q^2}{\tau^2} & 0 \\ 0 & 0 \end{bmatrix} \text{ and } R = r^2$$

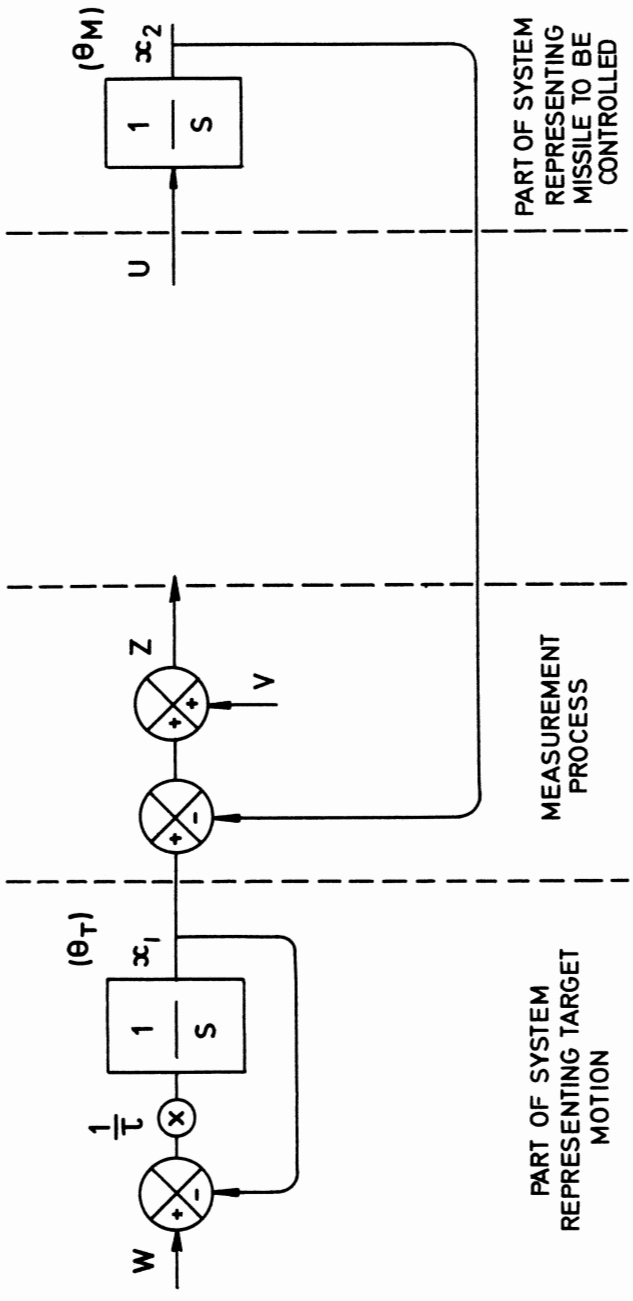


FIG 12.5-2 To obtain best estimates of x_1 and x_2

Substituting these expressions for Q and R, together with the expressions for A, B and H from equations (12.5-3) and (12.5-4), into the Riccati equation (B5) and expanding gives:

$$\dot{p}_{11} = -\frac{2p_{11}}{\tau} + \frac{q^2}{\tau^2} - \frac{1}{r^2}(p_{11} - p_{21})^2 \qquad (12.5\text{-}5a)$$

$$\dot{p}_{21} = -\frac{p_{21}}{\tau} - \frac{1}{r^2}(p_{21} - p_{22})(p_{11} - p_{21}) \qquad (12.5\text{-}5b)$$

$$\dot{p}_{22} = -\frac{1}{r^2}(p_{21} - p_{22})^2 \qquad (12.5\text{-}5c)$$

These can be solved for p_{11}, p_{21} and p_{22} by integrating the three simultaneous differential equations forward in time. To do this values for $p_{11}(0)$, $p_{21}(0)$ and $p_{22}(0)$ must be assumed. In the absence of any further information it is convenient to set $p_{11}(0) = p_{21}(0) = p_{22}(0) = 0$ and the results obtained when taking these initital conditions and particular values for q^2, r^2 and τ are shown in Figure 12.5-3.

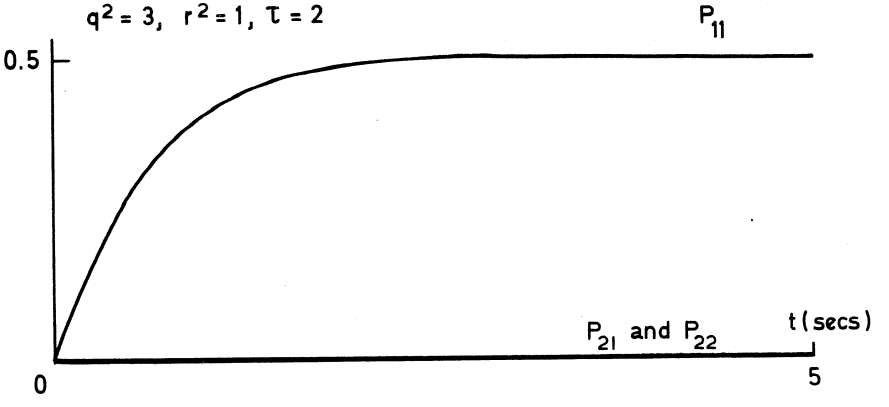


FIG 12.5-3 Solution of equations (12.5-5)

Since p_{12} and p_{22} appear to have zero steady state values we were lucky in our choice of initial conditions in that only $p_{11}(0)$ was given an incorrect value. It is pertinent to consider the effect of different choices for the initial conditions. This can be demonstrated by taking, say, $p_{11}(0) = p_{21}(0) = p_{22}(0) =$ 0.25 as shown in Figure 12.5-4.

In this case, convergence to the steady state values is taking much longer since we have started with three incorrect initial conditions.

Because the initial behaviour of the covariance matrix P is governed by the initial guesses for $p_{11}(0)$, $p_{21}(0)$ and $p_{22}(0)$ we will concentrate only on the

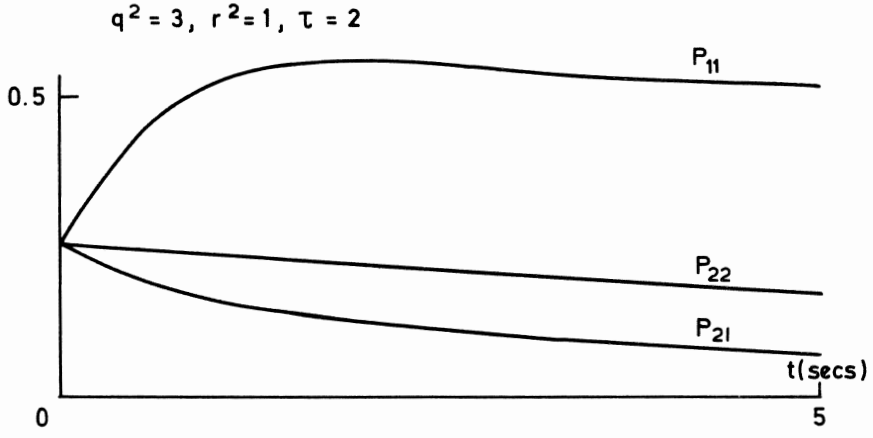


FIG 12.5-4 Solution of equations (12.5-5)

steady state situation, i.e. $\dot{P} = 0$. Although this can be studied by integrating forwards until a steady state condition is reached, as in Figure 12.5-3, such a procedure only gives results for specific values of q^2, r^2 and τ^2. However, by putting $\dot{p}_{11} = \dot{p}_{21} = \dot{p}_{22} = 0$ in equations (12.5-5) and solving the resulting algebraic equations, it is readily verified that

$$p_{21} = p_{22} = 0$$

$$\text{and } p_{11} = \frac{r^2}{\tau}\left(\sqrt{1 + \frac{q^2}{r^2}} - 1\right)$$

in the steady state.

Remembering that p_{11} and p_{22} are the respective variances of $\hat{x}_1$ and $\hat{x}_2$, the above results indicate that, even in the presence of measurement noise, the state x_2 can be estimated without error although some uncertainty will always persist in the estimate of x_1. This can be explained by looking again at the system shown in Figure 12.5-2. That part of the system representing the missile to be controlled, i.e. the state x_2, is known exactly. When this is modelled in the Kalman filter it can be used to estimate x_2 quite independently of the noisy measurement. However, the model for the target motion can never be perfect since the white noise w cannot be included within the Kalman filter. In fact, we can now form the complete Kalman filter for the steady state condition. We have

$$P = \begin{bmatrix} \frac{r^2}{\tau}\left(\sqrt{1 + \frac{q^2}{r^2}} - 1\right) & 0 \\ 0 & 0 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & -1 \end{bmatrix}, \quad R = r^2$$

Hence, from equation (B4), the Kalman filter gain matrix is

$$K = PH^T R^{-1} = \begin{bmatrix} K_{11} \\ K_{21} \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau}\left(\sqrt{1 + \frac{q^2}{r^2}} - 1\right) \\ 0 \end{bmatrix}$$

and the filter structure, deduced from equation (B3), is as shown in Figure 12.5-5. Comparing this with Figure 12.5-2 it is seen that we have now completed the task we set ourselves at the beginning of the chapter, viz how to combine an *a priori* system model with a noisy measurement to form best estimates of all the system states.

12.6 COMBINED FILTERS AND CONTROLLERS

In section 11.3 we tackled a stochastic control problem on the assumption that the states x_1 and x_2, representing target and missile positions, could be measured exactly and we proceeded as if we had a deterministic system by substituting a mean value, i.e. zero, for the white noise source.
What happens when all the state variables are not available and/or the measurements of them are contaminated with noise?
Clearly, we can use a Kalman filter to obtain best (minimum variance) estimates of all the states in this case. However, will we still have an optimal controller if we use these best estimates $\hat{x}_1$ and $\hat{x}_2$ instead of the true states in implementing the optimal control?
Fortunately, it can be shown that such is the case and this result is known as *the separation theorem* for reasons that should be apparent from Figure 12.6-1.

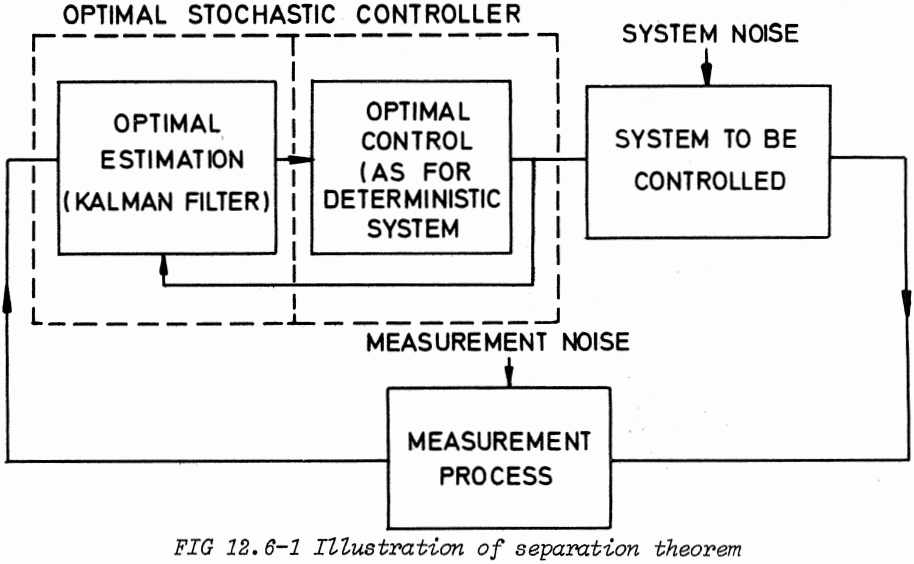


FIG 12.6-1 Illustration of separation theorem

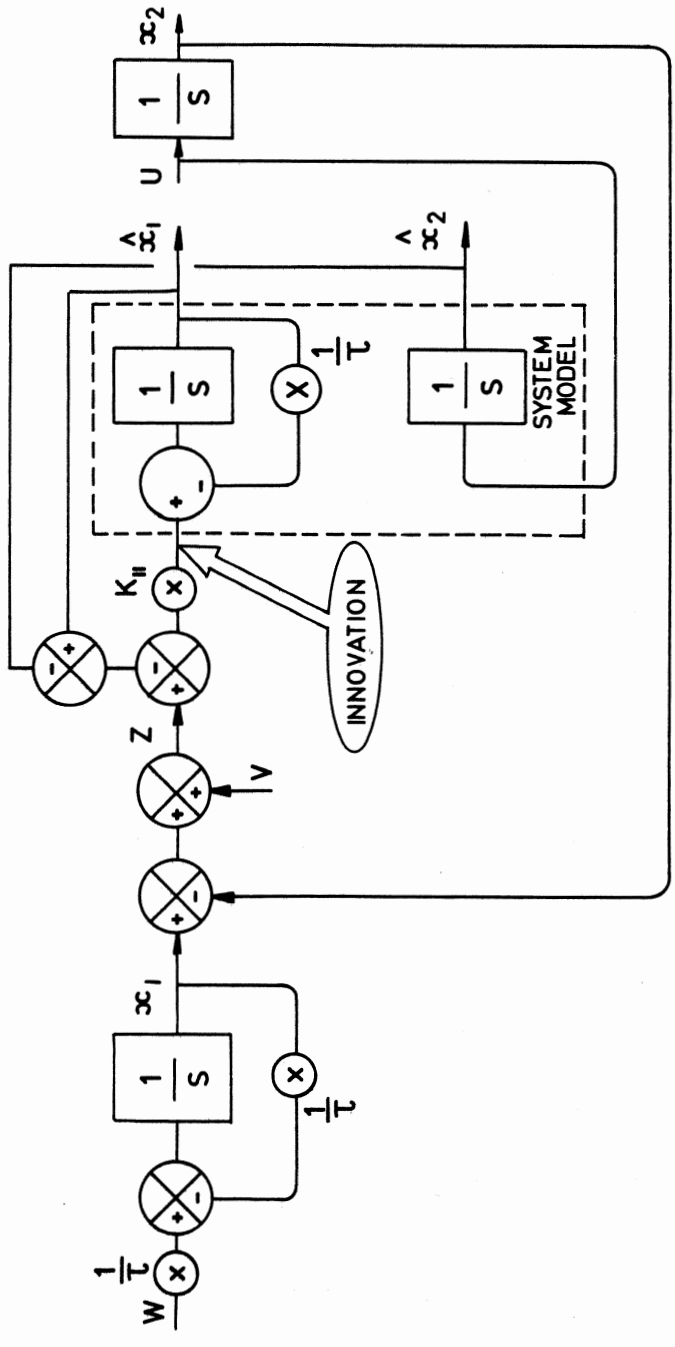


FIG 12.5-5 Kalman filter for two state system

In other words, we can compute separately the optimal control as if there is no noise and then use best estimates supplied by a Kalman filter in order to implement it. Engineers experienced in classical techniques will recognise this as a normal design procedure. It is merely because we have become so precise in specifying our control as "optimal" and the estimates as "minimum variance" that such theorems are necessary.

Hence, we can combine the steady state optimal controller derived in section 11.3 with the Kalman filter derived in section 12.5 and produce a combined filter/controller as shown in Figure 12.6-2.

More specifically, by taking $q^2 = 3$ and $r^2 = 1$, corresponding to the spectral densities of the system and measurement noises used in the original Wiener filter problem (*Example* 10.4-1), then

$$K_{11} = \frac{1}{\tau}\left(\sqrt{1 + \frac{q^2}{r^2}} - 1\right) = \frac{1}{\tau}$$

and it is readily shown that the closed loop transfer function relating θ_M to $(\theta_T + \theta_N)$ is given by

$$\frac{\theta_M}{(\theta_T + \theta_N)} = \frac{\frac{\tau}{\tau + \sqrt{\lambda}}}{(1 + \sqrt{\lambda}s)(2 + \tau s)}$$

and, in particular,

$$\frac{\hat{x}_1}{(\theta_T + \theta_N)} = \frac{1}{(2 + \tau s)}$$

This is just the result we have been expecting. The Kalman filter has contributed the extra factor necessary to take account of measurement noise and the control and estimation functions, derived separately in chapters 11 and 12 respectively, have combined to give the same overall system as that derived using Wiener filter methods in chapter 10. Further examples of combining optimal controllers and estimaters, particularly applied to CLOS systems, can be found in references (4), (5) and (6).

12.7 MODERN VERSUS CLASSICAL DESIGN

Provided the system we are dealing with is linear and the desired performance can be expressed in terms of a quadratic performance index then we have seen that the optimal control can be derived using a well defined procedure (appendix A). Furthermore, if any uncertainties in the system or measurements can be represented by zero mean white noise sources then a similar procedure is available to obtain best estimates of all the states in order to implement that optimal control (appendix B).

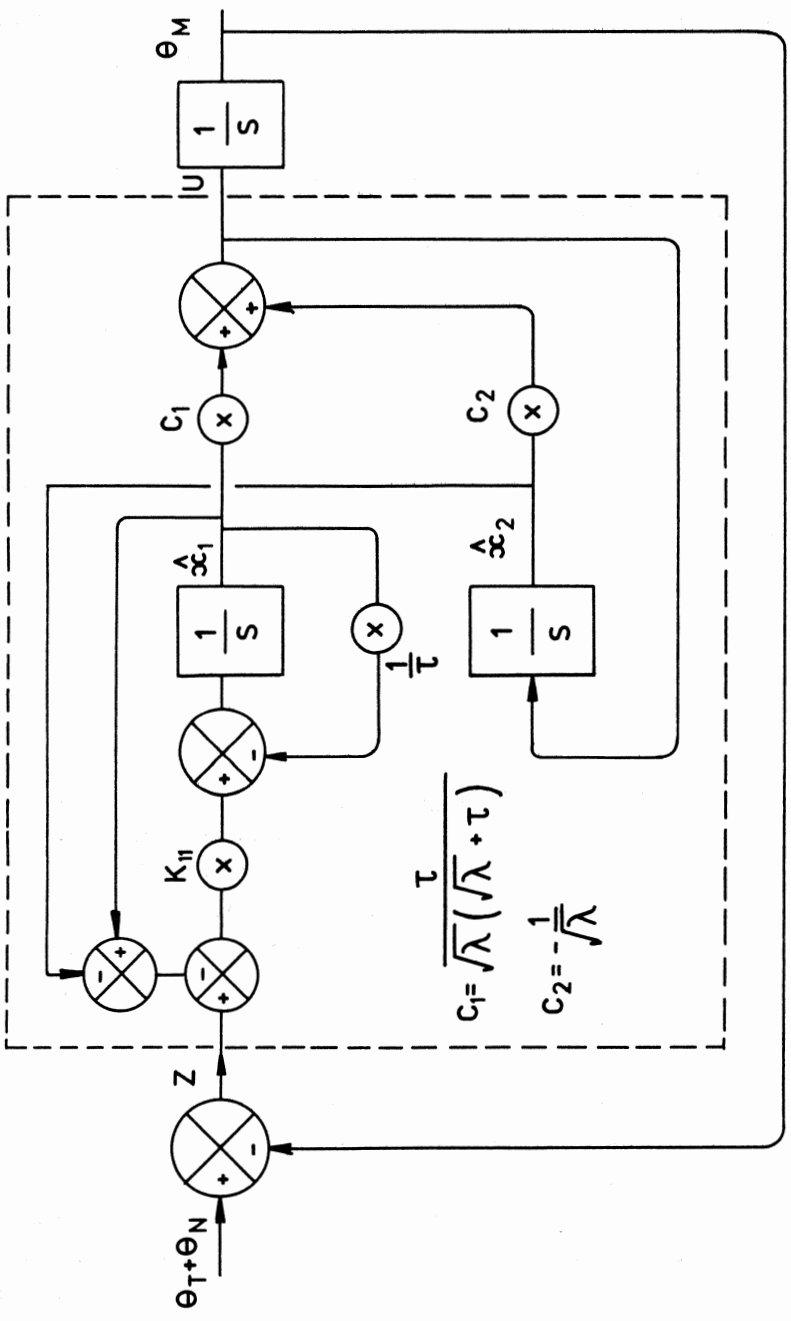


FIG 12.6-2 Combined filter/controller for two state system

It can be argued that the assumptions made in pursuing this approach are too unrealistic and therefore the results obtained are of academic rather than practical interest. However, all theoretical engineering techniques rely on simplification to some extent but they are justified provided they lead in the right direction to the final solution of the problem. We have seen that the modern approach leads to a well defined controller structure (e.g. Figure 12.6-2). Even if the parameter values are subsequently found to be unsatisfactory, as a result, say, of a simulation which includes some previously neglected effect, at least a basic design is available on which further development can be done.

The modern design will often seem more complicated than one arrived at by classical means. This is because all usable information concerning the system to be controlled is considered from the outset and this is immediately reflected in the controller which will have a system model embedded in it. In the classical frequency domain approach this same information is embodied in just one line on a Nyquist plot. Whilst this concentrates attention on the stability problem, it does so by excluding reference to the detailed system structure and a fairly simple controller usually results. As other aspects of the system design are considered, for example noise, extra terms are introduced into the controller and this procedure continues in an ad-hoc manner until a satisfactory design is achieved. In effect, more and more components of the original system structure are reintroduced into the controller. However, it is unlikely that the final design will be as complex as that arrived at directly by modern methods. The designer will have used his judgement and decided that any extra complexity would not lead to further significant improvements in performance. Similarly, had the designer started with the complete structure of a modern controller he would certainly consider simplifying it up to the point where a significant degradation in performance was noted. In both cases the inevitable engineering compromise would be made and, hopefully, it would result in very similar final solutions. Indeed, for many single-input single-output systems, the modern techniques do nothing more than provide a complementary approach to that offered by classical techniques. However, for cases where a particular control system structure is not apparent from previous experience or where multiple-inputs and multiple-outputs are being considered, the modern methods provide a very powerful framework on which to build. Furthermore, although we have concentrated only on linear systems, the state space approach, used in conjunction with a digital computer to solve the necessary equations, is the

natural technique to adopt when dealing with non-linear time-varying systems. In conclusion, it should be noted that although the application of modern control theory to *practical* situations is still in its infancy, it must be expected that increasing use will be made of these techniques by the guided weapon industry in the forthcoming years. Just as the computing power of the large digital machine has made possible the off-line optimisation of system control, so the recent introduction of the microcomputer has provided a further impetus towards the possibility of on-line optimisation, or true adaptive control, within the missile itself.

REFERENCES

1. BARHAM P.M. and D.E. HUMPHRIES Derivation of the Kalman filtering equations from elementary statistical principles. RAE Tech Report 69095 1969.

2. BRYSON A.E. and Y. HO Applied Optimal Control. Wiley 1975.

3. ANDERSON B.D.O. and J.B. MOORE Linear Optimal Control. Prentice-Hall 1971.

4. RICHARDS L.J. and M.R. ABBOTT A comparison of two methods of constructing the optimal transfer function for a simplified guided-weapon system. RAE Tech Memo MATH 7307 July 1973.

5. HEAP E. Combined Kalman filters and optimal control in analogue form for guided weapon systems. RAE Tech Memo WE 1386 March 1974.

6. HEAP E., P.J. HERBERT, H. LEWIS and T.E. COWLARD The equivalence of feedforward techniques, Wiener filters and steady-state Kalman filters in command-to-line-of-sight missile guidance design. RAE Tech Memo GW(NEW) 1014. July 1975.

APPENDIX A

OPTIMAL CONTROL OF LINEAR SYSTEM WITH QUADRATIC PI

APPENDIX A - Optimal control of linear system with quadratic PI

It is felt that it is more important to gain practice in the use of optimal control methods before pursuing a detailed justification of the techniques used. To this end the treatment given here is suggestive rather than thorough and more rigorous derivations can be found in most textbooks concerned directly with optimal control theory, e.g. reference 6 (ch 11).

It is recommended that the example of a one dimensional system considered in section 11.2 be studied before proceeding with this multidimensional treatment. Familiarity with the idea of the transpose of a matrix (denoted here by the superfix T) and with matrix manipulation is required.

Assume a multidimensional linear system represented by

$$\dot{\underline{x}} = A\underline{x} + B\underline{u} \tag{A1}$$

Our objective is to find a control vector $\underline{u}$ that will minimise the PI

$$J = \underline{x}^T P\underline{x}(T) + \int_{t_o}^{T} \{\underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}\}\, dt \tag{A2}$$

where P, Q and R are symmetric matrices. This restriction on the form of the matrices ensures that the PI retains its quadratic form in the multi-dimensional case. As the examples in chapter 11 indicate, this is not a very severe restriction in practice.

The term $\underline{x}^T P\underline{x}(T)$ in the PI allows some emphasis to be put on the minimisation of the final values of selected states.

Hence, following Pitman (1), we note that

$$\int_{t_o}^{T} \{\frac{d}{dt}\, \underline{x}^T P\underline{x}\}\, dt = \int_{t_o}^{T} \{\underline{x}^T \dot{P}\underline{x} + \dot{\underline{x}}^T P\underline{x} + \underline{x}^T P\dot{\underline{x}}\}$$

$$= \int_{t_o}^{T} \{\underline{x}^T \dot{P}\underline{x} + \dot{\underline{x}}^T P\underline{x} + \underline{x}^T P\dot{\underline{x}} + \underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}\}\, dt$$

$$- \int_{t_o}^{T} \{\underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}\}\, dt$$

(by first adding and then subtracting the term $\underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}$)

$$= \underline{x}^T P\underline{x}(T) - \underline{x}^T P\underline{x}(t_o) \tag{A3}$$

Hence $$J = \underline{x}^T P\underline{x}(T) + \int_{t_o}^{T} \{\underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}\}\, dt \quad \text{restating equation (A2)}$$

$$= \underline{x}^T P\underline{x}(t_o) + \int_{t_o}^{T} \{F\}\, dt \quad \text{from equation (A3)}$$

where $$F = \underline{x}^T \dot{P}\underline{x} + \dot{\underline{x}}^T P\underline{x} + \underline{x}^T P\dot{\underline{x}} + \underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}$$

Minimising both sides of this expression for J with respect to $\underline{u}$ gives

$$J_{opt} = \underline{x}^T P\underline{x}(t_o) + \min_{\text{wrt } \underline{u}} \int_{t_o}^{T} \{F\}\, dt$$

since $\underline{x}^T P\underline{x}(t_o)$ is dependent only on the starting time t_o and is unaffected by $\underline{u}$.

Hence the PI will be minimised provided a control vector $\underline{u}$ can be found that will make the integrand F equal to zero for $t_o < t < T$.

By substituting for $\dot{\underline{x}}$ from equation (A1) in the expression for F we have

$$F = \underline{x}^T \dot{P}\underline{x} + (A\underline{x} + B\underline{u})^T P\underline{x} + \underline{x}^T P(A\underline{x} + B\underline{u}) + \underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}$$

Now $$(A\underline{x} + B\underline{u})^T P\underline{x} = (A\underline{x})^T P\underline{x} + (B\underline{u})^T P\underline{x}$$

$$= \underline{x}^T A^T P\underline{x} + \underline{u}^T B^T P\underline{x}$$

Therefore

$$F = (\underline{x}^T \dot{P}\underline{x} + \underline{x}^T A^T P\underline{x} + \underline{x}^T PA\underline{x} + \underline{x}^T Q\underline{x}) + (\underline{u}^T B^T P\underline{x} + \underline{x}^T PB\underline{u}) + \underline{u}^T R\underline{u} \tag{A4}$$

This can be written in the form of a perfect square

$$F = (K\underline{x} + \underline{u})^T R(K\underline{x} + \underline{u}) \tag{A5}$$

$$= (\{K\underline{x}\}^T + \underline{u}^T)R(K\underline{x} + \underline{u})$$

$$= \underline{x}^T K^T R(K\underline{x} + \underline{u}) + \underline{u}^T R(K\underline{x} + \underline{u})$$

$$= (\underline{x}^T K^T RK\underline{x}) + (\underline{u}^T RK\underline{x} + \underline{x}^T K^T R\underline{u}) + \underline{u}^T R\underline{u} \tag{A6}$$

provided $$K^T RK = \dot{P} + A^T P + PA + Q \tag{A7}$$

and $$RK = B^T P \tag{A8}$$

(by comparing coefficients in equations (A4) and (A6)). If equations (A7) and (A8) are satisfied then the integrand is always non-negative and the minimum possible value (i.e. zero) of the integral term will be achieved provided

$$\underline{u} = \underline{u}_{opt} = -K\underline{x} \qquad \text{from equation (A5)}$$

$$= -R^{-1}B^T P\underline{x} \quad \text{substituting for } K \text{ from equation (A8)} \tag{A9}$$

Hence $\underline{u}_{opt}$ can be found provided P can be determined. Eliminating K from equations (A7) and (A8) gives

$$(R^{-1}B^TP)^TR(R^{-1}B^TP) = \dot{P} + A^TP + PA + Q$$

which, on simplifying, becomes

$$-\dot{P} = A^TP + PA + Q - PBR^{-1}B^TP \tag{A10}$$

Equation (A10) is known as the Matrix-Riccati equation which must be solved for P in order to determine the optimal control given by equation (A9).

In summary:

Given a system $\dot{\underline{x}} = A\underline{x} + B\underline{u}$ (A1)

and PI

$$J = \underline{x}^T P\underline{x}(T) + \int_{t_o}^{T} \{\underline{x}^T Q\underline{x} + \underline{u}^T R\underline{u}\}\, dt \tag{A2}$$

then the optimal control that minimises J is

$$\underline{u}_{opt} = -R^{-1}B^TP\underline{x} \tag{A9}$$

where P is derived from the Riccati equation

$$-\dot{P} = A^TP + PA + Q - PBR^{-1}B^TP \tag{A10}$$

APPENDIX B

OPTIMAL ESTIMATION — THE CONTINUOUS KALMAN FILTER

APPENDIX B - Optimal estimation - the continuous Kalman filter

Given a system

$$\dot{\underline{x}} = A\underline{x} + B\underline{u} + \underline{w} \tag{B1}$$

and measurement

$$\underline{z} = H\underline{x} + \underline{v} \tag{B2}$$

where $\underline{w}$, $\underline{v}$ are white noise vectors with spectral density matrices Q, R respectively, then the best estimate $\hat{\underline{x}}$ of $\underline{x}$ is given by

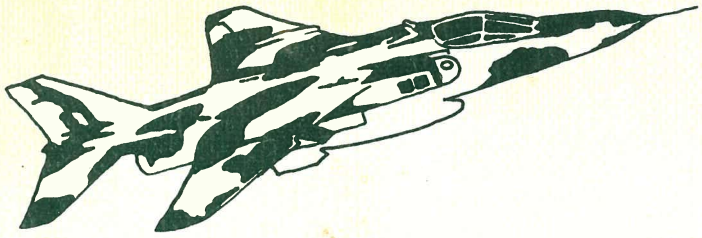
$$\dot{\hat{\underline{x}}} = A\hat{\underline{x}} + B\underline{u} + K\,(\underline{z} - H\hat{\underline{x}}) \tag{B3}$$

$$\text{where } K = PH^T R^{-1} \tag{B4}$$

and P is derived from the Riccati equation

$$\dot{P} = AP + PA^T + Q - PH^T R^{-1} HP \tag{B5}$$

INDEX



Guided Weapon Control Systems

P. Garnell & D.J. East

CONTENTS

0 08 019691 8